OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: October 28, 2001, 17:17:26 ; Search time 3262.14 Seconds
                 (without alignments)
                 11763.899 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgcttgctcgct..........agaaggtcaagagccagtga 2481

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
                 Listing first 45 summaries

Database :   GenEmbl:*

1:	gb_ba1:*	44:	em_ov:*
2:	gb_ba2:*	45:	em_pat:*
3:	gb_ba3:*	46:	em_ph:*
4:	gb_in1:*	47:	em_pl:*
5:	gb_in2:*	48:	em_ro:*
6:	gb_in3:*	49:	em_sts:*
7:	gb_om:*	50:	em_sy:*
8:	gb_ov:*	51:	em_un:*
9:	gb_pat1:*	52:	em_vi:*
10:	gb_pat2:*	53:	gb_sts1:*
11:	gb_ph:*	54:	gb_sts2:*
12:	gb_pl1:*	55:	gb_sts3:*
13:	gb_pl2:*	56:	gb_sy:*
14:	gb_pl3:*	57:	gb_un:*
15:	gb_pl4:*	58:	gb_vi1:*
16:	gb_ba1:*	59:	gb_vi2:*
17:	gb_ba2:*	60:	gb_htg1:*
18:	em_fun:*	61:	gb_htg2:*
19:	em_htgo_hum:*	62:	gb_htg3:*
20:	em_htgo_inv:*	63:	gb_htg4:*
21:	em_htgo_rod:*	64:	gb_htg5:*
22:	em_htg_hum1:*	65:	gb_htg6:*
23:	em_htg_hum2:*	66:	gb_htg7:*
24:	em_htg_hum3:*	67:	gb_htg8:*
25:	em_htg_hum4:*	68:	gb_htg9:*
26:	em_htg_hum5:*	69:	gb_htg10:*
27:	em_htg_hum6:*	70:	gb_htg11:*
28:	em_htg_hum7:*	71:	gb_htg12:*
29:	em_htg_hum8:*	72:	gb_htg13:*
30:	em_htg_inv1:*	73:	gb_htg14:*
31:	em_htg_inv2:*	74:	gb_htg15:*
32:	em_htg_other:*	75:	gb_htg16:*
33:	em_htg_rod:*	76:	gb_htg17:*
34:	em_hum1:*	77:	gb_htg18:*
35:	em_hum2:*	78:	gb_htg19:*
36:	em_hum3:*	79:	gb_htg20:*
37:	em_hum4:*	80:	gb_htg21:*
38:	em_hum5:*	81:	gb_htg22:*
39:	em_hum6:*	82:	gb_htg23:*
40:	em_hum7:*	83:	gb_htg24:*
41:	em_in:*	84:	gb_htg25:*
42:	em_om:*	85:	gb_prl:*
43:	em_or:*	86:	gb_pr2:*
		87:	gb_pr3:*
		88:	gb_pr4:*
		89:	gb_pr5:*
		90:	gb_pr6:*
		91:	gb_pr7:*
		92:	gb_pr8:*
		93:	gb_pr9:*
		94:	gb_ro1:*
		95:	gb_ro2:*
		96:	gb_in4:*
		97:	gb_pr10:*
		98:	em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2481	100.0	2908	89	AF304370	AF304370 Homo sapi
2	2477.8	99.9	2997	91	BC001939	BC001939 Homo sapi
3	2477.8	99.9	3006	91	BC004158	BC004158 Homo sapi
4	2455.4	99.0	2908	89	AF308698	AF308698 Pan trogl
5	2448.2	98.7	2976	89	AK001392	AK001392 Homo sapi
6	2442.6	98.5	2893	89	AF308694	AF308694 Gorilla g
7	1686.2	68.0	2712	94	AF308696	AF308696 Mus muscu
8	247.4	10.0	118788	85	AC005277	AC005277 Homo sapi

```
RESULT   1
         AF304370
LOCUS       AF304370     2908 bp    mRNA            PRI       23-FEB-2001
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            (HPC2/ELAC2) mRNA, complete cds.
ACCESSION   AF304370
VERSION     AF304370.1  GI:10880932
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED    11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     CDS             1..2481
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24441.1"
                     /db_xref="GI:10880933"
                     /translation="MWALCSILRSAAGRIMSQGRRISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVLQVVAAGSRDSGAALIVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKEFGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IEALVRPHSAPEYEDETMTVYQIPIHSEORGRKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNPLVLRAKEMGLPVGTHAIAPIIA
                     AVKDGKSITHEGRETILAEELCTPDFQGAAPVVVCEGSEFQIPCENEAFQRYQSKAD
                     APVALVVHHAPASVLVDSRXQOWMERFGPDTQHLVLNENCASVHNLRSHKIOTQLNI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPREWQRDAIITCNPEEFIVEA
                     LQLENFQOSVOEYRSRSAODGPAPAEKRSOYPEIIFIGLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCCEBGTFGQLCRIHGDQVDRVLGTLAAVPVSHLHADHHTGLPSILLQRRAL
                     ASLGKPLHPLLVVAPNQLRAWLQOHNQCOEVLHHISMIPAKCLOEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKAVVSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSOAISVGMRMNAEFIMLNHFSORYAKVPLFSPNFS
                     EKVGVAFDHNKVCFGDFPTMPKLIPPLKALFAGDIEEMEMEREKRELRQVRAALLSRE
                     LAGGLEDGEPQOKRAHTEEPQAKKVRAQ"
BASE COUNT      702 a    793 c    826 g    587 t
ORIGIN
```

Sequence hit summary:

No.	Score				ID	Accession	Organism
9	146.2	5.9	740	89	AF304371S2	AF304369	Homo sapi
10	144.4	5.8	34593	94	AF348157	AF348157	Mus muscu
11	133.4	5.4	2690	5	AF215894	AF215894	Drosophil
12	133.4	5.4	115873	64	AC017383	AC017383	Drosophil
13	133.4	5.4	175682	4	AC007417	AC007417	Drosophil
14	133.4	5.4	190574	60	AC007352	AC007352	Drosophil
15	133.4	5.4	261846	5	AE003830	AE003830	Drosophil
16	124	5.0	721	89	AF304371S1	AF304371	Homo sapi
17	84.2	3.4	31201	15	SPAC1D4	Z69939	S.pombe chr
18	71.8	2.9	46335	12	AC022354	AC022354	Arabidops
19	70	2.8	19158	13	AF188714	AF188714	Emericell
20	67.8	2.7	952	53	CNS079WW	AL435846	T7 end of
21	67	2.6	32329	15	SPBC3D6	Z95620	S.pombe chr
22	65	2.6	143209	13	AP001550	AP001550	Oryza sat
23	58.8	2.4	75289	12	AB023046	AB023046	Arabidops
24	58.4	2.4	3685	1	Z28304	Z28304	S.cerevisia
25	56.8	2.3	23533	1	AE001039	AE001039	Archaeogl
26	53.8	2.2	99241	88	AC079030	AC079030	Homo sapi
27	53.8	2.2	145440	69	AC026119	AC026119	Homo sapi
28	53.8	2.2	181289	73	AC068797	AC068797	Homo sapi
29	52.4	2.1	12026	1	AE005109	AE005109	Halobacte
30	51.8	2.1	951	53	CNS06JSP	CNS06JSP	S.pombe
31	51.8	2.1	951	53	CNS06HXX	AL399595	T3 end of
32	51.8	2.1	11836	1	AE000936	AE000936	Methanoba
33	51.4	2.1	729	97	HSI18896	Y18896	Homo sapien
34	51.4	2.1	982	97	HSY18894	Y18894	Homo sapien
35	51.4	2.1	1086	91	HOSAL7573	Y17573	Homo sapien
36	51	2.1	3215	59	HS2VP16B	H2SVP16B	Herpes simp
37	50.8	2.0	33779	3	SCGB3	AL095822	Streptomy
38	50.6	2.0	14189	93	HSPLECTIN	Z54367	H.sapiens g
39	50.6	2.0	22693	97	HSU53204	U53204	Human plect
40	50.6	2.0	22693	93	HSPLEC1S3	U63610	Human plect
41	49.8	2.0	105795	3	SYCCPNC	D64001	Synechocyst
42	49.2	2.0	226013	74	SYCCPNC	AC073680	Mus muscu
43	49.2	2.0	226013	74	AC073680	AC073680	Mus muscu
44	49	2.0	92858	69	AC024959	AC024959	Homo sapi
45	48.6	2.0	39150	3	MTCY21B4	Z80108	Mycobacteri
45	48.2	1.9	171	59	HS4IR3NA	K00106	epstein-bar

```
                             ALIGNMENTS

         Query Match        100.0%;  Score 2481;  DB 89;  Length 2908;
         Best Local Similarity  100.0%;
         Matches 2481;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 atgtggggcgctcttgctgtctgggttcggggtccggggtccggaggacgcacagtgcaggaggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTGGGGCGCTCTTGCTGTCTGGGTTCGGGGTCCGGGGTCCGGAGGACGCACAGTGCAGGAGGC 60

Qy     61 accatatcgcaggcaccgcaccgccccgccccgcgcagcggccaaggaccccgctgccggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCATATCGCAGGCACCGCACCGCCCCGCCCCGCGCAGCGGCCAAGGACCCCGCTGCCGGCACCTG 120

Qy    121 cgcacgcgagagaagccgggaccgtcggggttgctccggcgcccaaacacgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCACGCGAGAGAAGCCGGGACCGTCGGGGTTGCTCCGGCGCCCAAACACGTGTACCTG 180

Qy    181 caggtggttggcagcgggggtagccgggactcgggggtccggggtgcttccggagttc 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGGTTGGCAGCGGGGGTAGCCGGGACTCGGGGGTCCGGGGTGCTTCCGGAGTTC 240

Qy    241 aaccggtatctccttcaactgtggagaaggcagtcatgcaggagcacaagtta 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AACCGGTATCTCCTTCAACTGTGGAGAAGGCAGTCATGCAGGAGCACAAGTTA 300

Qy    301 aaggtgctcgcctggacaacatattcctgacacgaatgcactggtctaatgtggggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTGGGGGGC 360

Qy    361 ttaagtggaatgatctcttacttaaggagaaacggggcttcaaaagtgtgtacttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTAAGTGGAATGATTCTTACTTAAGGAGAAAACCGGGCTTCAAAAGTGTGTACTTCTGGA 420

Qy    421 cctccacaactggaaaaataccctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTCCACAACTGGAAAAATACCCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 480
```

```
Qy    481 atagaactggctgtgcggcgcccaactcctgcccagaatacgagggatgaaacctgacagtt  540
          ||||||||||||||||||||||| ||||||||||||||||||||||||||| ||||||||||
Db    481 ATAGAACTGGCTGTGCGGCGCCCACTCTGCCCAGAATACGAGGGATGAAACCATGACAGTT  540

Qy    541 taccagatcccacacagtgaacacagtcccaccatggcagagtcca                  600
          |||||||||||| ||||||||||||||||| |||||||||||||||
Db    541 TACCAGATCCCATACACAGTGAACACAGTGCCACCATGGCAGAGTCCA                600

Qy    601 gaaaggccctcagcaggtcagtcccagacgatcccagtcgaatgaaat                660
          ||||||||||| |||||||||||||||| ||||||||||||||    |
Db    601 GAAAGGCCCTCTCAGGTCAGTCCCATGAACAGCACCAACCATGGCAGAGTCCA          660

Qy    661 gagccacaccttccacatggtgttagccagggaaggggtcaggggactcttccctggtc    720
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db    661 GAGCCACACCTTCCACATGGTGTTAGCCAGGGAAGGGGTCAGGGGACTCTTCCCTGGTC    720

Qy    721 gtagctttcatcttgaagcttcacttaaagagagagagaagaaag                   780
          ||||||||||||||||||| |||||||||| |||||||||||||||
Db    721 GTAGCTTTCATCTTGAAGCTTCACTTAAACAGAGAGAGAAGAAAG                   780

Qy    781 gagatgggcctccagagttgggacacagctgcactcgctcccatccagaaggac         840
          ||||||||||||| |||||||||||||||||||| ||||||||||||||||||||
Db    781 GAGATGGGCCTCCCAGAGTTGGGACACAGCTGCACTCGCTCCCATCATTGCTGTCAAAGGAC 840

Qy    841 gggaaaagcatcactcatgaaggaagagagagatttggctgaaagagagagagaagaaag   900
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    841 GGGAAAAGCATCACTCATGAAGGAAGGAGATTTTGGCTGAAAGAGAGTGTTACTCCTCCA    900

Qy    901 gatcctggttgctgttggttagaatgtccagatgaaagctccattcattgctgctttcctcca 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GATCCCTGGTCGTCTGCTTTTGTGGTAGAATGTGTCAGATGAAAGCTTCAATCCCATC     960

Qy    961 tgtgtgaagaatgccacactttcagaggtacccaagacatcctggtgttactcctca      1020
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    961 TGTGAGAATGCCACACTTTCAGAGGTACCCAAGGAAAAGCAGATTTTTGGCTGGATGAGG   1020

Qy   1021 gttcacatggcccccagccagcaatctgtgcttgtggacaagcaagtgacctggtgg      1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCCCAGCCATCTGTTGCTTGTTGGACACAGCAAGTACCAGTGGAATGGAGG 1080

Qy   1081 tttgggccctgaacacccagccactgtttgttgggctgagaatgtccaggtcaaccctt   1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGGCTTGAACACCTGAACCCAGCACTTGGTCCTGAATGACAAGCTTCATTCACAACCTT 1140

Qy   1141 cgcagccacaagattcaaaccgagagagtcaacctcatcccacccggaacttcttccccgctc 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAGCCACAAGATTCAAACCCAGTCTTCAACCTCTATCCCACGACAACTTCACCCCTGCCTC 1200

Qy   1201 accagttccgctgtaagagaggagggccccacccagtgtgcccatggttcaagggtgaa    1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAGTTTCCGCTGTAAGAGAGGAGGGCCCACCTCAGTGTGCCCATGGTTCAAGGGTGAA    1260

Qy   1261 tgcctcctcaagtacagctccgtcccagaggggaggaggatgcattattact           1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGCCTCTCAAGTACCACAGCTCCGTCCCAGAGGGGAGGGGATGCCATTATTACT           1320

Qy   1321 tgcaatcctgaggaattcattagttgaggcgctgcaggttcccaactccagcagaggtg    1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGCAATCCTGAGGAATTCATAAGTTGAGGCCCTCAGCAGAGGACGGGCTCCCCGCAGGCCA   1380

Qy   1381 caggagtacaggaggagtgcgcgcaggacgccccagaagaaaaagaagtcagtgac        1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGGAGTACAGGAGGAGTGCGGCAGGACGCCCCAGAGAGAAAAAGAAGTCAGTAC        1440

Qy   1441 ccagaaatcatcttccttggaacaggggtctgccatccgatgaagattcgaaatgtcagt   1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCATGGTCTCTGCTGAGAAGAGATT  1500

Qy   1501 gccacacttgtcaacataaggccccgacaacgtctctgctactggactgtgtcagggaggacaca 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCACACTTGTCAAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGTGGTTCAGGGGCACA 1560
```

```
Qy   1561 tttggggcagctgtgcctgtcattacggaggacagggtgacaggggtcctgtgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGGGCAGCTGTGCCGTCCGTCATTACGGAGGACCAGGTTGCCAAGTATCTTGCTG       1620

Qy   1621 gctgtgtttgtttgtcccaacctgcacgcaggatcaccacaggggcttgccaagtactttgctg  1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCTGTGTTTGTGTCCCAACCTGCACGCAGGATCACCACAGGGGCTTGCCAAGTACTTTGCTG   1680

Qy   1681 caggagaacgcggccttggcatcttttgggaagcgcttcaccctttgctgtggttgcc       1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGGAGAGAACGCGGCCTTGGCATCTTTTGGGAAGCGCTTCACCCTTTGCTGTGGTTGCC    1740

Qy   1741 cccaaccagctcaaagccttggcctcccaacagcagtaccacacaaccagtgccaaggagg    1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCAAACCAGCTCAAAGCCTGGCTCCAGAGTACCACACAACCAGTGCCAAGGAGG         1800

Qy   1801 cacatcagtatgattcctgccaaatgtccttcaggaggaaggggctgagagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACATCAGTATGATTCCTGCCAAATGTCCTTCAGAGGAAGGGGCTGAGAGATCTCCAGTCCTGCA 1860

Qy   1861 gtgggaaagattgatcagttcgctgttggcgaacatgctgtttgggaacatgtag         1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTGGGAAAGATTGATCAGTTCGCTGCCAAATGCCTTCAGAAAGTGGGAAACATGTGT       1920

Qy   1921 ctggtgcggcactgcaagcatgcgtttgctgtcggtgttggcggtgcacactctgctggaaa     1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTCGGTGCACCTTTCGGTGCGGGAA         1980

Qy   1981 gtggtctattccggggacaccagtctgggaacatgtccctcctggggaggaaagatgcc       2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GTGGTCTATTCCGGGGACACCAGTCTGGGAACATGTCCCCTCCTGGGGAGGAAAGATGCC   2040

Qy   2041 acccttcctgaatacatgaagacccaagtcgcgtggaaggtgttggactctgagacactcgct 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ACCCTTCCTGAATACATGAAGACCCAAGTCGCTGTTGGAAGGTGTTGGAGAGGCACAGT    2100

Qy   2101 acacacagcacaacgtcccaacagccaatcagcgtggggatgcgggagttcatt          2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACACACAGCACAACGTCCCACTGAACAGCCAATCATCAGCGTGGGGATCGGGAGTTCATT    2160

Qy   2161 atgctgaccactccagccaacgctatgccaaggtcccctcctcctgtttcaacctcagc      2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ATGCTGAACCACTCCAGCCAACGCTATGCCAAGGTCCCCCTCCTGTTCAGCTTCATTCAGC  2220

Qy   2221 gagaaagtggggagttgccttttgaccacacatgaaagtctgcgtttgaccaacaatg      2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGAAAGTGGGGAGTTGCCTTTTGACCACACATGAAAGTCTGCGTTTTGACCAACAATG    2280

Qy   2281 cccaagctgattccccccactgaaagccctgttctgtgggcgactcgaggagatggaggag   2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CCCAAGCTCATTCCCCCACTGAAAGCCCTGTTCTGTGGGCGACACGAGGGGAGGATGGAG   2340

Qy   2341 cgcaggggagaagtgcgcgcaggactgccaggtcgcaggagagatcgacgagatgcggga    2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGCAGGGGAGAAGCGCGCAGGACTGCAGGTTCGCAGGAGAGATCGGACATGCGGGA       2400

Qy   2401 ggcgggccctggaggatgggaagcctcagagaaggaggagcacagggcc              2460
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GGCGGGCCCTGGAGGATGGGGAAGCCTTCAGCAGGATGGGGAAGCCCCACAGGCC        2460

Qy   2461 aagaaggtcagagcccagtga                                          2481
          |||||||||||||||||||||
Db   2461 AAGAAGGTCAGAGCCCAGTGA                                          2481
```

RESULT 2
BC001939

```
LOCUS       BC001939                2997 bp    mRNA            PRI       16-MAR-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:4102, mRNA, complete cds.
ACCESSION   BC001939
VERSION     BC001939.1  GI:12804972
KEYWORDS    MGC.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2997)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: 1 Column: 17.
FEATURES             Location/Qualifiers
     source          1..2997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:4102 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             61..2541
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH01939.1"
                     /db_xref="GI:12804973"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVILQVVAAGSRDSSAALYVFSENRYLENCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLGSMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVVQIPIHSEQRGKHQPWQSPERPLSRLSPERSSDSSELN
                     ENEPHLPHGVSQRGRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTRAAIAPIIA
                     AVKDKSITHEGREILAEELCTPPDEGAAFVVWECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSPRCKKEGPTLSVPMVQGECLLKYQLRPREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSAFLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKRPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPARCLQEGAEISSPAVERL
                     ISSLRTCDLEEFQTCLVRHCKHAFGCCALVHTSGWRKVVYSGCDTMPCEALVRMGKDATL
                     IHEARLEDGIEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRIAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDFPTMFKLIPLKALFAGDIEMEEEREKRELRQVRAALLSRE
                     LAGGLEDGEPQOKRAHTEEPQAAKKYRAQ"
BASE COUNT     740 a    806 c    853 g    598 t
ORIGIN
```

Query Match   99.9%; Score 2477.8; DB 91; Length 2997;
Best Local Similarity   99.9%;
Matches 2479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgtgggcgcttgctcgctgcgggtccgggccggacgaccaccatgtcgcaggagacgc   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATGTGGGCGCCTTTGCTCGCTGCGGTCGCGGCGCTGGACGACGACCATGTCGCAGGAGAGC  120

Qy     61 accatatcgcaggagacccgccgcccgcgaggcgcgcagcggaggaccgcggcaccgcactg  120
Db    121 ACCATAATCGCAGGACACCCGCCGCCCGCGAGGCGCGCAGCGGAGGAGCCGCGGCACCGCACTG  180

Qy    121 cgcacgccggagaagcgcggaaccgtcggg  ...  180
Db    181 CGCACCCGCGACGAGAACCGCGGAACCGTCGGG  ...  240

Qy    181 cagg ...  240
Db    241 CAGG ...  300

Qy    241 accggtatctcttcaaactgtggagaaggcgttcaagaactcatgcaagagcacaagtta  300
Db    301 AACCGGTATCTCTTCAAACTGTGGAGAAGGCGTTCAAGACTCATGCAGGAGCACAAGTTA  360

Qy    301 aaggttgctccgcctgggacaacatattcctcgacacagaatgcactggttctaatggggc  360
Db    361 AAGGTTGCTCCGCCTGGGACAACATATTCCTCGACACAGAATGCACTGGTTGTGGGGGC  420

Qy    361 ttaagtggaatgattctttaaagggaaaccgggcttccacaaagttgttgtactttctgga  420
Db    421 TTAAGTGGAATGATTCTTTAAAGGGAAAACCGGGCTTCCACAAAGTGTGTTACTTTCTGGA  480

Qy    421 cctccacaactgaaaaatacctcgaagcaatcaaaatatttctggttccattgaaagga  480
Db    481 CCTCCACAACTGGAAAAAATACCTCGAAGCAATCAAAATATTTTCGGTTCCATTGAAAGGA  540

Qy    481 atagaactggcttgtcgggccccactctgccccagaaatacgagggatgaaaccattaacagtt  540
Db    541 ATAGAACTGGCTGTGTCGGCCCCCACTCTGCCCCAGAAATACGAGGATGAAACCATGACAGTT  600

Qy    541 taccagatcccatacacacagttgacacagaggaggggaaagcgcaccaacatggcagagtcca  600
Db    601 TACCAGAATCCCATACACACAGTGGACACAGAGGAGGGGAAAGCACCAACATGGCAGAGTCCA  660

Qy    601 gaaaggcctctccagcaggctcagtgacacgatctcagacctccgagtcgaatgaaaat  660
Db    661 GAAAGGCCTCTCCAGCAGGCTCAGTGACACGATCTTCAGACTCCGAGTTGAATGAAAAT  720

Qy    661 gagccacaccttccacaatggtgtgttagcccagaagaaggaagggttcagggactcttcctggtc  720
Db    721 GAGCCACACCTTCCACAATGGTGTTAGCCCAGAAGAAGGAGGGGTCAGGGACTCTTCCTGGTC  780

Qy    721 gtagctttcatctgtaaagcttcaactcactaaagaagagaaaatcttctggttctcaaagcaaag  780
Db    781 GTAGCTTTCATTCGTGTAAAGCTTCACTTAAGAAGAGAAAAACTTCTTGGGTGTGCTCAAAGCAAAG  840

Qy    781 gagatgggcctcctcccagctggagacagctgccatcgtcccacatcattgctgtctgtcaaggac  840
Db    841 GAGATGGGGCCCTCCCCAGTTGGGACAGCTGCCATCGTCCCACATCATTGCTGTGTCAAGGAC  900

Qy    841 gggaaaagcatcactcatgaaggagaagaatttggctgaaagagctgtgtactcctcca  900
Db    901 GGGAAAAAGCATCACTCATGAAGGAGAGAGAATTTGGCTGAAGAGCTGTGTTACTCCTCCA  960

Qy    901 gatcctggtgctgctttgtggttggtagaatgtccaagtgaaagcttcattcaaccctc  960
Db    961 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC  1020

Qy    961 tgtgagaatgcacacctttcaggaggtaccaaggaaaggcagatgccccgtggcttggtc  1020
Db   1021 TGTGAGAATGCACACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCGTGGCCTTGGTG  1080

Qy   1021 gttcacatggcccagcaatctgttgctcttgggacagcagcagtggatggaggagg  1080
Db   1081 GTTCACATGGCCCCAGCAATCTGTTGCTCTTGGGACAGCAGCAGTGGATGGAGGAGG  1140

Qy   1081 tttggggcctgacaccagacacttggttcctgaatgagaacttcacaaactt  1140
Db   1141 TTTGGGGCCTGACACCCAGACACTTGGTTCTGAATGAGAACTGTTGCCTCAGTTCACAAACCTT  1200
```

```
Qy  1141 cgcagcccacaaagattcaaaaccagctcaacctcatccaccccgacatctccccctgctc 1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CGCACCCACAAAGATTCAAAACCAGCTCAACCTCATCCACCGGACATCTCCCCCTGCTC 1200

Qy  1201 accagtttccgtctgtaagaaggagaggcccccaccccttcaagtgtgccaggtcaggtgaa 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 ACCAGTTTCCGCTGTAAGAAGGAGAGGGCCCCCACCCTTCAGTGTGCCCATGGGTTCAGGGTGAA 1320

Qy  1261 tgcctcctcaaagtaccagctcccgtcccaagagggaggggatgccattattact 1320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TGCCTCCTCAAGTACCAGCTCCCGTCCCAAGGAGGGGATGCCATTATTACT 1380

Qy  1321 tgcaatcctgaggaattcatcatagttgaaggcgtgaagcttcccaacttccagcagagcgtg 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 TGCAATCCTGAGGAATTCATCATAGTTGAAGGCGTCCGGTCCCAGGAGGGACTGGCACAGGGC 1440

Qy  1381 caggagtacaaggaggagtgcgcaggacagcgcccccaccgaggacgagaagaagtcagtac 1440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CAGGAGTACAAGGAGGAGTGCGCAGGACAGCGCCCCCACCGAGGAGGAGAAGAAGTCAGTAC 1500

Qy  1441 ccagaaaatcatcttccttggaacaggggtctgccatccccgatgaagaattcgaaatgtcagt 1500
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 CCAGAAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGAATTCGAAATGTCAGT 1560

Qy  1501 gccacaacttgtcaacataagccccgaacacgtctctgctactggagctgtggtgaggcaca 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GCCACACTTGTCAACATAAGCCCCGAACAGTCTCTGCTACTGGACTGTGTGGTTGAGGGCACG 1620

Qy  1561 tttgggcaggtgtgccgtcattacggaggaccaggtggacaggtcctggcaaccctggct 1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 TTTGGGCAGGTGTGCCGTCATTACGGAGGACCAGGTGGACAGGTCCTGGCACCCTGGCT 1680

Qy  1621 gctgtgtttgtgtgtccccacctgcacgcagcagatcaccacacacgggcttgccaagtatcttgctg 1680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACACGGGCTTGCCAAGTATCTTGCTG 1740

Qy  1681 cagagagaacgcgccttggcaatcttttggaaaagccgcttcaaccctttgctggttggttgcc 1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 CAGAGAGAACGCGCCTTGGCAATCTTTGGGAAAAGCCGCTTCAACCTTTGCTGGTTGGTTGCC 1800

Qy  1741 cccaaccagctcaaagcctggcctcacagcagtaccacaacaagtgccaggaggtcctgcac 1800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 CCCAACCAGCTCAAAGCCTGGCCTCCAGCAGTACCACAAACAAGTGGCTCCAGGAGGTCTGCAC 1860

Qy  1801 cacatcagtatgattcctgccaaatgcctcaggaaggggctgagatctccagtcctgca 1860
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGCTGAGATCTCCAGTCCTGCA 1920

Qy  1861 gtggaaagattgatcagttcgctgttgcgaacatgtgattggaagagttcagacctgt 1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTGGAAGAGTTTCAGACCTGT 1980

Qy  1921 ctggtgcggaactgcaagcatgctgtttgctggtgcacacctctcggctggaa 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 CTGGTGCGGAACTGCAAGCATGCGTTTGGCTGTGCGCCTGGTGCACACCTCTGGCTGGAAA 2040

Qy  1981 gtggtctattccggggacacatgcccctcggctcggtccggatgtgggggaaagatgcc 2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 GTGGTCTATTCCGGGGACACCATGCCCTCGGCTCTGGTCCGGATGTGGGGGAAAGATGCC 2100

Qy  2041 accctcctgatacatgaagccacctggaagatggttggaagaagacagtgaaaag 2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGACAGAGGAAAAAG 2160

Qy  2101 acacacagcacacaaagtcaacaaagcgcggatgaacgcggagttcatt 2160
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 ACACACAGCACACAACGTCCCAAGCCATCAGCGGTGGGGATGAACGCGGGAGTTCATT 2220

Qy  2161 atgctgaaccacttcagccaagcgctatgccaaggtccccctcttcaagcccaaacttcagc 2220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 ATGCTGAACCACTTCAGCCAAGCGCTATGCCAAGGTCCCCTCTTCAAGCCCAAACTTCAGC 2280

Qy  2221 gagaaagtggggagttgcctttaccacacatgaagttgtcttgcttttggagaactttcaaacaatg 2280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 GAGAAAGTGGGAGTTGCCTTTGACCACACATGAAGGGTCTGCGTTTTGGAGACTTTCCAACAATG 2340

Qy  2281 cccaagctgattcccccaactgaaagccctgttctgcgcgacatcgaggaggagatggaggag 2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 CCCAAGCTGGAATTCCCCCACTGAAAGCCCCTGTTTGCTGGCACATCGAGGAGGAGATGGAGGAG 2400

Qy  2341 cgcagggggaagcgcgggggaccagtgcgggcaggtgcggggcggggcccctcctgtccaggggacctggca 2400
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 CGCAGGGGAGGAAGCGGGGGACCAGTGCGGGCCAGGTGCGGGCGGGGCCCTTCGTCCTGTCCAGGGGACTGGCA 2460

Qy  2401 ggcggggcctggaggatgggggagcctcagcagaggaagcgggggcccacacagagaggaggagccacaggcc 2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GGCGGCCCTTGGGAGGATGGGGAGCCTCAGCAGAGAGGCCCCACACAGAGGAGCCACAGGCC 2520

Qy  2461 aaggaaggtcaagagcccagtga 2481
         |||||||||||||||||||||||
Db  2521 AAGGAAGGTCAGAGCCCAGTGA 2541
```

```
ORIGIN

BASE COUNT    742 a    808 c    858 g    598 t

Query Match     99.9%;     Score 2477.8;     DB 91;     Length 3006;
Best Local Similarity   99.9%;   Pred. No.: 0;
Matches 2479;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;
```

ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
IELAVRRHSAPEYEDETMTVVQIPIHSEQRGKHQPWQSPERPLSRLSPERSSDSELN
ENEPHLPHGVSQRGRVRDSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
AVRDGKSITHEGREILAEELCTPDPGAAFVVVECDEDSFIQPICENATFQRYQGKAD
APVALVVHMAPASVLVDSRYQQWMERGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPREWRQRDAIITCNPEEFIVEA
LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPIFIFLGTGSAIPMKIHRNVSATLVNISP
DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
ISSLLRICDLLEFQTCLVRHCKHAFGCALVHHSGMKWVVYSGDTMPCEALVRMGKDATL
LIHEATLEDLGEEEAVEKTHSTTSQAISVGMRMNAEEIMLNIFSQRVAKVPLFSPNFS
EKVGVAFDHMKVCFGDPPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"

```
Qy      1 atgtgtgggcgcttgctctcgctgcgtctccgggggcccggagccggccggccg   60
          |||| ||||| ||||| |||||||||||||||||||||||||||||||||||||
Db     70 ATGTGGGCGCGCTTTGCTCTCGCTGCGTCTCCGGGGCCCGGAGCCGGCCGGCC  129

Qy     61 accatatcgcaggcaccgccccgccccgcccgcgcaagggacccgtcgcacctg  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 ACCATATCGCAGGCACCGCCCCGCCCCGCCCGCGCAAGGGACCCGTCGCACCTG  189

Qy    121 cgcacgcgagaagcgcggaccgtcgggttgtctccgggcggccggcctgtacctg  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 CGCACGCGAGAAGCGCGGACCGTCGGGTTGTCTCCGGGCGGCCGGCCTGTACCTG  249

Qy    181 caggtcgtggtgcaggtgagtcggggactcggggtctcggggcccagttc  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 CAGGTCGTGGTGCAGGGGTAGCGGGGACTCGGGGTCTCCGCGCCCAGTTC  309

Qy    241 aaccggtatctcttcaactgtggagaagcgttcaggagactcatgcaagta  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 AACCGGTATCTCTTCAACTGTGGAGAAGCGTTCAGGAGACTCATGCAAGTTA  369

Qy    301 aaggttgctcgcctggacacaatatcctgacacgtcactgttctaatgtgggggc  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 AAGGTTGCTCGCCTGGACACAATATTCCTGACACGTCACTGTTCTAATGTGGGGC  429

Qy    361 ttaagtggaatgaattcttactttaaaggaaaccgggcttccaaagtgtgtacttc  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 TTAAGTGGAATGAATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTCGGA  489

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaaatattctgtgtccattgaaagga  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAAATATTTTCTGTGTTCATTGAAAGGA  549

Qy    481 atagaactgctgtgtcggcccactctgcccagaattacgaggatgaaacctgacagtt  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 ATAGAACTGGCTGTGTCGGCCCCACTCTGCCCCAGAATTACGAGGATGAAACCATGACAGTT  609

Qy    541 taccagaatcccataccacacagtgaacagagaggggaaagcacaacagaagtcca  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    610 TACCAGAATCCCCATACACACAGTGAACAGAGAGGGGAAAGCACAACCATGGCAGAGTCCA  669

Qy    601 gaaaggcctcagcagcgtcagtccagagctccgagtcgaatgaaat  660
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    670 GAAAGGCCTCCAGCAGCGCATCTTCAGAGCTCCGAGTTGAATGAAAAT  729

Qy    661 gagcacacacacttccacatggttgttagccagagagggttcagggaactcttccctggtc  720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    730 GAGCCACACCTTCCACATGGTTGTTAGCCAGAGAGGGTTCAGGGACTCTTCCCTGGTC  789

Qy    721 gtagcttcactgtaagcttcacttaagggtgctcaaaagcaaag  780
          |||||||||||||||||||||||||||||||||||||||||||
Db    790 GTAGCTTTCATCTGTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTCCTCAAAAGCAAAG  849
```

```
Qy    781 gagatgggcgtcccagttgggacagctgccatcgctgtccccatcaatgctggctgtcaaggac  840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    850 GAGATGGGCGCCTCCCAGTTTGGGACAGCTGCCATCGCTGTCCCCATCAATTGCTGTCAAGGAC  909

Qy    841 ggggaaaagcatcaccatgaaggagagaagaattctggctgaagagctgtgtgtactcctca  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    910 GGGGAAAAGCATCACCTCATGAAGGAGAGAGAGAATTTTGGCTGAAGAGACTGTGTGTACTCCTCCA  969

Qy    901 gatcctggtgcttgctcttgtggctgggtagaagtgtccagaatggcgaaagcttcatccaaccac  960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    970 GATCCTGGTGCGTCGCTTTTGTGGGTGGGTGGTAGAATGTCCAGAATGGCGAAAAGCTTCAACCCATC  1029

Qy    961 tgtgagaagtgccacactcttcagaggttaccaaaggacagcagagcagccccgtggccttggtg  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1030 TGTGAGAAATGCCACACTTTTCAGAGGTTACCAAAGGACAGCAGAGTGCCCCGTGGCCTTGGTG  1089

Qy   1021 gttcaacatggccccaacatctgtgcttgtggacacagcagccagagtgggcttgacacctt  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1090 GTTCACATGGCCCCAACCATCTGTGCTTGTGGACACAGCCAGCATGTGGGACACTTCACAACCTT  1149

Qy   1081 tttgggcctgacacccacagcacttggtcctgaatgaatgaagacagcgtctc  1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1150 TTTGGGCCTGACACCCACAGCACTTGGTCCTGAATGAATGAAGACAGCGTCTC  1209

Qy   1141 cgcgcaaagaattcaaaaccaacaatgcagagtctggacactggggtcctcccctgtc  1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1210 CGCAGCACAAGAATTCAAAACCCCAGCTCAAACCTCATCACCCCCGGACACTGGTCCTCAGCTC  1269

Qy   1201 accagtttccgctgtaaggaggagggccccacccctcagtgttccaatggttcaggttgaa  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1270 ACCAGTTTTCCGCTGTGAAGGAGGGGCCCACCCCTCAGTGTGTCCAGTGGTTCCAGGTTGAA  1329

Qy   1261 tgcctcctcaagtaccagctccgtcccagcaagctccgagcaggaggaggggatgccattact  1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1330 TGCCTCCTCAAGTACCAGCTCCGTCCCAGCAAGCTCCGAGCAGGAGGAGGGGATGCCATTATTACT  1389

Qy   1321 tgcaatcctgaggaattcatagttgaagtgctgcaagcttcccaacttcagcaaacgtg  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1390 TGCAATCCTGAGGAATTCATAGTTGAGGTGCTGCAAGCTTCCCAACTTCCCCAACTTCAGCAGACG  1449

Qy   1381 caggaggtacagggagttgcgcaggcccagccccaacagagaaaaaaagtcaagtac  1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1450 CAGGAGGTACAGGGAGTTGCGCAGGACGGGCCCCAGCCCCAACAGAGAAAAAAAGTCAGTAC  1509

Qy   1441 ccagaaatcatctccttgggacacggtctgcccatcccgatgaagaatgtcagt  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1510 CCAGAAATCATCTCCTTGGGACACGGGTCTGCCCATCCCCGATGAAGAATTGCAAATGTCAGT  1569

Qy   1501 gcccacacttgtcaacattaagcccgacacgtctctgctactggcaagtgttgtgaggcaca  1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1570 GCCCACACTTGTCAACATTAAGCCCCGACACGTCTCTGCTACTGGCAAGTGTTGTGGAGGGCACG  1629

Qy   1561 tttgggcagctgtgccgtcattacgtagacgagggacaaggtgggacacgggcact  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1630 TTTGGGCAGGCTGTGTGCCGTCATTACCGGGACCAGGGGTGGGACAGGGTCCGGGCT  1689

Qy   1621 gctgtgttttggtgtcccacgtcacgcagaatcaacaacaggggcttggccaagtactttgctg  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1690 GCTGTGTTTGTGTCCCACCTGCACGCAGACCATCACACAACGGGGCTTGGCCAAGTATCTTGCC  1749

Qy   1681 caggagacaacgggccttgccatcttcacccttttgtttttgggattggctgtcc  1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1750 CAGAGAGAACCGGGCCTTGCCATCTTTGGACATCTTTGCACCCTTTTGCTGGTGGGTTTGCC  1809

Qy   1741 cccaaccagtcaaagctggctccagacagtaccacacacccagagagttctggac  1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1810 CCCAACCAGTCAAAGCTGGCTCCAGACAGTACCACAACAACCAGAGAGGTCTGGAC  1869

Qy   1801 cacatcaatgatccctgccaaatgcctcagaggatcctccagaggtgccagga  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1870 CACATCAGTGTATGATGATTCCTGCCAAATGCCTTCACTGAGAGATCTCCAGAGGTCCTGCA  1929

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgattgtttggaagagagttcagaccgt  1920
```

```
RESULT 4
LOCUS       AF308698    2908 bp    mRNA        PRI    27-FEB-2001
DEFINITION  Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION   AF308698
VERSION     AF308698.1   GI:10946496
KEYWORDS    .
SOURCE      chimpanzee
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED    11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
```
```
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     CDS             1..2481
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24920.1"
                     /db_xref="GI:10946497"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPAREERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGTLQRLMQEHKLKV
                     ARLDNIFLITRMHWSNVGGLSGMIITLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPYEDETMTVYQIPIHSEQRGRGKHQPWQSPERLSRLSPERSSDSESN
                     ENEPHLPHGVSQRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGRRILAEELCTPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPESVLVDSRYQQWMEREGPDTLHVLRNENCASVHNLRSHKIQTQLNLI
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTSSAIPMKIRNYSATLVNISP
                     DTSLLLDGCEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLNLILLORERAL
                     ASLGKPFFIPLLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEBGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKIHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQALSVGMRMNNAEFIMLNHSORYAKVPLSPNFN
                     EKVGVAFDHMKVCFGDFATMPKIIPPLAVAIFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEPQAKKVRAQ"
BASE COUNT      712 a     788 c     819 g     589 t
ORIGIN
```

```
Qy       1 atgtggggcgcgttctcgctctgcgtctgcggttccggcggccgggacgacgcaccatgtcgcaagggacgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGTGGGGCGCGTTTCGCTCTGCGTCTGCGGTTCCGGCGGCCGGGACGACGCACCATGTCGCAGGGACGC 60

Qy      61 accatatcgcaaggcaccccgccccgccccgagcgcgcaaggagaccccgctgcgccacctg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACCATATCGCAGGCACCCCGAGGCACCCCCGCCCCGAGCGCGCAAGGAGACCCCGCTGCGCCACCTG 120

Qy     121 cgcacgcgagaagcgcggaccgtcgggttgctcccggcggccaaaacaccgtgtacctg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGCACGCGAGAAGCGCGGACCGTCGGGTTGCTCCCGGCGGCCAAAACACCGTGTACCTG 180

Qy     181 caggtggtggcaggggtagcccggaactgggcgcgccgcgtctacgtctccgagttc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CAGGTGGTGGCAGGGGTAGCCCGGGACTCGGGCGCCGCGCTCTACGTTCTCCGAGTTC 240

Qy     241 aaccggttgctcttcaactgtggagaaggcgttcagagacactcatgcaggagcacaagtta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AACCGGTACCTCTTCAACTGTGGAGAAGGCCATTCAGAGACTCATGCAGGAGCACAAGTTA 300

Qy     301 aaggttgctcgctggaccaacatattcctgaccacagaatgacactggtctaatgttggggc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAGGTTGCTCGCCTGGACAACATATTCCTGACAACACAGATGCACTGGTCTAATGTTGGGGGC 360
```

Then the alignment portion (upper section) continues, paired Qy/Db:

```
Db    1930 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTCATTGGAAGAGTTCAGACCTGT 1989
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1921 ctggtgccggcactgacatgcgtgttggctgtcgctggttgcacactctgggaa 1980

Db    1990 CTGGTGCCGCAAGCACTGCACGCATCCGTTTGGCTGTCGCCGGTTGCACACTCTGGAA 1989

Qy    1981 gttcggtctattccggggacaccatgccctcgggcctgcctcgtccggtagtgggaaagatgcc 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2050 GTGGGTCTATTCCGGGGACACCATGCCCTGCCCTGCGTCCGGATGGGAAGATGCC 2049

Qy    2041 acccctcctgatacatgaagcacccctggaagaatggttggaggaagaagcagtggaaag 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2110 ACCCTCCTGATACATGAAGCACCCCTGGAAGAATGGTTTGGAGGAAGAAGCAGTGGAAAAG 2109

Qy    2101 acacacaagcaacgtcccccaagccatcagcgtgtgggatgggagtgaacgcggagttcatt 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2170 ACACACAAGCAACACAACGTCCCAAGCCATCAGCGTGGGGATGGGAACGCGGAGTTCATT 2169

Qy    2161 atgctgaaccacttcagccagcgctatgccaaggtcccccctcttcagccccccaacttcagc 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2230 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGC 2229

Qy    2221 gagaaagtggggagttgcctttgaccacatagaaggttctgcttttgagagaactttccaacaatg 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2290 GAGAAAGTGGGGAGTTGCCTTTGACCACATAGAAGGTTCTGCTTTTGAGAGAACTTTCCAACAATG 2289

Qy    2281 cccaagctgattccccactgaaaacccctgtttgctgcgaactcgaggagatggaggag 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2350 CCCAACCTGATTCCCCACTGAAAAGCCCTGTTTGCTGCGGACATCGAGGAGATGGAGGAG 2349

Qy    2341 cgcaggagaagcgggagctgcaggtgtcgggcggccggcctcccctgttccaggagctggca 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2410 CGCAGGAGAAGCGGGAGCTGCAGGTGTCGGGCAGGAGCTGCGGGGAGCTGGCA 2400

Qy    2401 ggcggcctggaggatgggggaggcccctcagcgagaaggacggggccccacacaggcc 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2470 GGCGGCCTGGAGGATGGGGAGGCCCTCAGCAGAGAAGCGGGCCGCGCCACAGGCC 2469

Qy    2461 aagaaggtcagagcgcccagtga 2481
           |||||||||||||||||||||||
Db    2530 AAGAAGGTCAGAGCCCAGTGA 2550
```

```
Qy    361 ttaagtggaatgatcttacttaaagaagaaccgggcttccaaagtgtgtactttctgga 420
          |||||||||||||||||||||||||| |||| |||||||||| |||||||||||||||||
Db    361 TTAAGTGGAATGATCTTACTTTAAAGAAGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 420

Qy    421 cctcccacaactggaaaaatacctcgaaagcaatcaaaaattatttctgttccattgaaagga 480
          ||||||||| ||||||||||||||| ||||||||||||||||||||||| ||||||||||||
Db    421 CCTCCACAACTGGAAAAATACCTCGAAAGCAATCAAAAATTATTTTCTGGTCCATTGAAAGGA 480

Qy    481 atagaaactggctgtgtcggtgcccacctctgcccccagaaatacgaagggatgaacagtt 540
          |||||||||||||||| ||||||||||||||||||||||||| |||||||||||||||||
Db    481 ATAGAACTGGCTGTGCGGTGCCCACTCTGCCCCCAGAAATACGAGGGATGAAACCATGACAGTT 540

Qy    541 taccagatcccataccaacagtgaacagggctcagccgaactcttcagacttcagcatcca 600
          ||||||||| |||||||||||||||||||||||||||||| ||| ||||||||||||||||
Db    541 TACCAGATCCCCATTACACAGTGAACACAGGCTCAGCCGGAAAACACCATGACAGTCCA 600

Qy    601 gaaaaggggcctcagcaggctcagtctccagagacctcagactcgaatgaaat 660
          |||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    601 GAAAGGGGGCCTTCACCAGCTGTCCAGAGCCGATCTTCAGACTTCACAGTCAATGAAAAT 660

Qy    661 gagccacaccttcccacatggtgttagccagagaagaagagagagagagagtcca 720
          |||||||||||||||||||| ||| |||||||| |||||||||||||||||||
Db    661 GAGCCACACTTCCCACATGGTGTTAGCCAGGAGAAAGAGAGAGAGAGAAGTCGGGGACTCAGCGAGTCC 720

Qy    721 gtagcttcaatcgtaagcttcacttaaagagagaaacttggtggctcaaggac 780
          ||| |||| ||||||| |||||||||||||||| ||||||||| ||||||||
Db    721 GTAGCTTTCATCGTGTAAGCTTCACTTAAAGAGAGAAAACTTGGTGGCTCAAAGCAAAG 780

Qy    781 gagatgggctcccagttgggacagctgccatcgccatccatcattgtgctgtcaaggac 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGATGGGCTCCCAGTTGGGACAGCTGCCATCGCCATCATTGCTCTGGTGTGTGTGGAC 840

Qy    841 gggaaaagccatcactcatgaagagaagaagaaattttggctgaaagagcttgtgtactcctcca 900
          ||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||||||
Db    841 GGGAAAAGCCATCACTCATGAAGAGAAGAAGAGATTTTGGCTGAAGAGCTGTGTTACTCCTCCA 900

Qy    901 gatcctggttgctgcttttgtggtagaagcttcattcaaacctt 960
          ||||||||||||| |||||||||||||||||||| |||||||||
Db    901 GATCCTGGTTGCTGCTTTTTGTGGTAGAAAGTCCAGATGAGAATCTCAGGGAAAGATGCC 960

Qy    961 tgtgagaatgccaacctttcagaggtggaccagagctgcctgccccgcgcctggtg 1020
          |||||||||||||||||||||||||||||||||| |||||||||||||||| ||||
Db    961 TGTGAGAATGCCACCTTTCAGAGGTACCAGAGCTGCCTGCCCCGTGGTGTGTGGGCCTTGGTG 1020

Qy   1021 gttcacatggcccagatctgtgcttggtggtaggtaccagcaagtgggatggga gg 1080
          |||||||| |||||| || |||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCAGAATCTGTGCTTGTTGGTACCAGCAAGTGGGATGGGAGG 1080

Qy   1081 tttgggcctgacaccacagcacttggtcttgatgtgcctcagttcacaaacctt 1140
          |||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db   1081 TTTGGGCAGTCGTGTTGCCCTCCATCGGAAGGAGTGCGAGGGATGCCATTATTACT 1140

Qy   1141 cgcagcacaagattcaaacccagctcaacctcatccatccctgctc 1200
          ||||||||||||||||| |||||||||||||||||||||||||||||
Db   1141 CCGAGCCACAAGATTCAAACCCAGCTCAACCTCATCACCGCTTGCTC 1200

Qy   1201 accagtttcccgtctgtaagagaggacccaccctcagtgtcgtgtgttcaggg 1260
          |||||||||||| |||| |||||||||||||||||||||| ||||||||||||
Db   1201 ACCAGTTTCCCCGTGTAAGGAGGGGCCCCACCTCAGTGTCGTGCCATTGGTTCAGGGTGAA 1260

Qy   1261 tgcctcctcaagtacccagctcccctcccaggagaggaggtgcaggagagtccat 1320
          ||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db   1261 TGCCTCTCTCAAGTTACCACCTCCCTGCGAGGGATGCCATTATTACT 1320

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagagcgtg 1380
          ||||||||||||||||||||| |||||||||||| |||||||
Db   1321 TGCAATCCTGAGGAATTCATAATTGAGGCGCTGCAGAGCGTGTGG 1380

Qy   1381 caggagtacagaggagtgcgcaggagagcccagccagtcac 1440
          ||||||||||||||||||||||| |||||||||||||||||
Db   1381 CAGGAGTACAGGAGGAGTGCGCAGGAGAGCGCCAGCGCAGAGAAAAGAGAGTCAGTAC 1440
```

```
Qy   1441 ccagaaatcatcttcctttgaacaagggtctgccatcccgatgaagaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1441 CCAGAAATCATCTTCCTTTGGAACAGGGTTCTGCTCTGCTACTGGCTACTCCGATGAAGAAGATTCGAAATGTCAGT 1500

Qy   1501 gccacacttgtcaaacataagccccgacacgtctctgctacttgctactcc 1560
          |||||||||||||||||||||||||||||||||||| |||||||||||||
Db   1501 GCCACACTTGTCAAACATAAGCCCCGACACGTCTCTGCTACTGGCACTGGTGTGGGCACG 1560

Qy   1561 tttgggcagctgtgccgtcattacggaccacct 1620
          |||||||||||||||||||||||||||||||||
Db   1561 TTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGGTCCTGGGCCACCTGGCT 1620

Qy   1621 gctgtcgtttgtccaacctgcacacgacgcagagcaagatcacacacgggccttgccaagtat 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCTGTGTTTTGTGGTCCGTCAAACCTGCACATGACCAGATCACAGTCACAGTCACAGGTCCTGGT 1680

Qy   1681 cagagagaaacgcgcctgagcaatcttttgggatcttttgggaaaagccttgccggtggttgccc 1740
          ||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGAGAGAAACGCGCCTGAGCAATCTTTTGGGATCTTTGGGAAAAGCCTTGCCGGTGGTTGCC 1740

Qy   1741 cacatcaagtatgattcagttgcgtggctctgcgaacaatgattcagagtcct 1800
          |||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db   1741 CACACAGTATGATTCCTGCCAAAGTGCTCCAGCAGTACCACAGTGGCAGTCCCTGCGAA 1800

Qy   1801 cacatcaagtatgattccctgcaaagtgctccagcagtaccacagtggcagtccctgcgaa 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACATCAGTATGATTCCTGCCAAAGTGCTCCAGCAGTACCACAGTGGCAGTCCCTGCGAA 1860

Qy   1861 gttggaaaggattgatcaagttcggtcgaacacattgaattcgaagacctgt 1920
          ||||||||||||| ||||||||||||||||||||||||||||| |||||||
Db   1861 GTTGGAAAGATTGATCAGTTCGGTGGTGTTTGCGAACATCGGAAGAGATTCCAGAGTTCCTGT 1920

Qy   1921 ctggtgcgggcaactgcaagcacatgcggttgggctggtgtgtgtgctgtggacccttggtg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1921 CTGGTGCGGGCACTGCAAGCACATGCGGTTGGGCTGGTGTGTGTGCACACCTCGGCTGGGAA 1980

Qy   1981 gtggtctattccgggggacacacattgcctccgaaggtgccctggagg 2040
          |||||| |||||||| ||| ||| |||||||||||||||||||||||
Db   1981 GTGGTCTATTCCGGGACACACAATGCCTCAGGTGCCGAGAATCTCAGTGAAAGATGCC 2040

Qy   2041 acccctcctgtacaatgaagtggccttgaccagaaatggttctgcctttggg 2100
          ||||||||||| |||||||||||| ||||||||||||||||||||||||||
Db   2041 ACCCTCCTGTACAATGAAGTGGCCTTTTGGAACGGTTTGGAACTTTGCACAAATG 2100

Qy   2101 acacacgacaagcacgctcccaagcccactgaacgtgggatggggatgaacgcggagttcatt 2160
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   2101 ACACACGACAAGCACGCTCCCAAGCCCACTGAACGTGGGATGGGAGTTCATT 2160

Qy   2161 atgctggaaccacacttcagccacggcgctatgccaaggccccaacttcagc 2220
          |||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   2161 ATGCTGGAACCACACTTCAGCCACGGCTATGCCAAGGTCCCAACTTCAAC 2220

Qy   2221 gagaaagtggagagtgtctgcttgtgacatagaagttccacaatg 2280
          |||||||||||||||||||||||| ||| |||||||||||||||
Db   2221 GAGAAAGTGGGGAGTTGCCTTTGACCACATGAAAGGTCTGCTTTGGAAGACTTTTGGGACAAATG 2280

Qy   2281 cccagctgattcccccactgaaagccctgttcttggcgacaatcgagaagaggag 2340
          ||||||||||| |||||||||||||||||||||||||||||||||||||||| |||
Db   2281 CCCCAAGCTGATTCCCACTGAAAGCCCTGTTCTTGGCGACAATCGGGAGAGGAG 2340

Qy   2341 cgcaggggaggaggctgcaggggggcggcttcctcgtttcgggacatcgagaatgaggag 2400
          ||||||||||| |||||||||||| ||||||||| ||||| |||||||||||||||||
Db   2341 CGCAGGGGAGAAGCGGAGCTGCAGGGGGGCGGCTTCCTCGTGTCCGGACATCGGAGAATGAGGAGTGGGAGA 2400

Qy   2401 ggcggcctgaggatcggggaaccctcagcagaaaccagcaggcc 2460
          |||||||||||||||||||| |||||||||||||||||||||||
Db   2401 GGCGGCCTGAGGATGGGGGAGCCCTCAGCAGAAAACCACCAGGCC 2460

Qy   2461 aaggagtacagaggagggcagacccagtga 2481
          |||||||||||||||||||||||||||| |
Db   2461 AAGAAGGTTCAGAGCCCAGTGA 2481
```

```
RESULT 5
LOCUS       AK001392        2976 bp     mRNA                     PRI        22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION   AK001392
VERSION     AK001392.1  GI:7022621
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2000985.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished (2000)
REFERENCE   2  (bases 1 to 2976)
AUTHORS     Isogai,T. and Otsuki,T.
TITLE       Direct Submission
JOURNAL     Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; Research Association for Biotechnology; cDNA library
            construction; 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RP2000985"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pME18SFL3; mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             75..2639
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91666.1"
                     /db_xref="GI:7022622"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGPNTVLQVVAAGSRDSGAALYVFSENRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMIILTKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAMRPHSAPEYEDETMTVYQIFIHSEQRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVFAFICKLHLKRGNFLVLRAKEMGLPVGTNAIAPTIA
                     AVKDGKSITHEGRIILAEELCTPDPDGAAFVVVECPDESFIQPICENATFQRYGCKAD
                     APVALVVHNAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTOLNLI
                     HDIFPLLTSFRCKEGPTLSVPMVQGECLLKYQLRPREWQRDAITTCNPEEFIVEAL
                     QLPNFQQSVQEYRSAQDGPAPAEKRSQYPEIIELGTGSAIPMKIRNVSATLVNISPD
                     TSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERALA
                     SLGKPLHPLLVAAPNQLKAWLQQYYNQCQEVLHHISMIPARCLQEGAEISSPAVERLI
                     SSLLRTCDLEEPQICLVRICKHAFGCALYHTSGWKVVYSGCDTMPCEALVRMGKDATLL
                     IHEAVLEGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSGRYAKAVPLFSPNFSE
                     KVGVAFPDHMVCLGDPFTMRKLIPPLKALFAGDIEMEEREKRELRQVRAALLSREL
                     AGGLEDGSLSRSGPTQRSHRPRRSEPSEDLGDPELRRLCVFCPTHAPVSALLAGRS"
BASE COUNT    712 a    807 c    856 g    601 t
ORIGIN
```

```
Query Match         98.7%;  Score 2448.2;  DB 89;  Length 2976;
Best Local Similarity   99.7%;  Pred. No. 0;
```

```
Matches 2474;  Conservative  0;  Mismatches  3;  Indels  4;  Gaps  2;

Qy          1 atgtggggcttgctcggctgcggtcgcggctcgcggccggacgcaccatgtcgcaggacgc 60
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db         75 ATGTGGGGCTTGCTCGGCTGCGGTCGCGGCTCGCGGCCGGACGCACCATGTCGCAGGACGC 134

Qy         61 accatatcgcaggcaccgcgcccgcccgcccgcagcgcgccgcaagcgcccgctcgcaccg 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 ACCATATCGCAGGCACCGCGCCCGCCCGCCCGCAGCGCGCCGCAAGCGCCCGCTCGCACCG 194

Qy        121 cgcacgcgagaaagcgcggaacgtcgggttgctccgcgggggccaaacccgtgtacctg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 CGCACGCGGAGAAAGCGCGGAACGTCGGGTTGCTCCGCGGGGGCCAAACACCGTGTACCTG 254

Qy        181 caggtggtcgcagcggtgagcgcggagcccggggctccacgtctctccgagttc 240
              ||||||||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db        255 CAGGTGGTGGCAGCCGGTGAGCGCGGGACTCGGGGCGCCCTCTACGTGTTCTCCGAGTTC 314

Qy        241 aaccggtatctcttcaactgtggagaagcgcgttcagagagactcatgcaggaggcacaagtta 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 374

Qy        301 aaggttgctccgctggacaacatattcctgacacgaatgcactggtctcaaatgttggggc 360
              ||||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        375 AAGGTTGCTCCGCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 434

Qy        361 ttaagtggaatgacttcacttaaagggaaacccgggcttccaaaagttgtgctacttctgga 420
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 TTAAGTGGAATGAATGATTCTTACTTTAAAGGAAAACCGGGCTTCCAAAAGTGTGTTACTTTCTGGA 494

Qy        421 cctccacaactggaaaaataccctcgaagcaatcaaaatatttctggtccatgaaagga 480
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        495 CCTCCACAACTGGAAAAAATAACCTCGAAGCAATCAAAATATTTTCTGGTCGAATGAAAAT 554

Qy        481 ataggaaactggctgtgcggcgccccactctgccccagaataacgaggagagcactgacagtt 540
              ||||||||||||||||| |||||||||||||| |||||||||||||||||||||||||
Db        555 ATAGAACTGGCTATGCGGCGCCCCACTCTGCCCCAGAATACGAGGGATGAAAACCATGACAGTT 614

Qy        541 taccagatccccatacacagtgaacagagaggaggaaacttcttggttgctcaaagcaaag 600
              ||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||||
Db        615 TACCAGATCCCCATACACAGTGAACAGAGAGGAGGAAACTTCTTGGTTGCTCAAAGCAAAG 674

Qy        601 gaaaggcctctcagcaggctcagtcaggcgatctcagcagcatctttcagacctcggaatgaaaat 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCTGCCATCCCATCATTCGTGTCGTCAAGGAC 734

Qy        661 gagccacaccttccacatgggtgttagccagagaagagaggggtcagggactctttccctggtc 720
              ||||||||||||||||| |||||| |||||||||||||||||||||||||||||||||
Db        735 GAGCCACACCTTCCACATCGGGTTGTTAGCCAGAGAAGAGAGGGGTCAGGGGACTCTTTCCCTCCA 794

Qy        721 gtagcttcaatcgtaaggcttcacttaaagagaggaaacttcttggttgctcaaagcaaag 780
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        795 GTAGCCTTTCATCGTAAGGCTTCACTTAAAGAGAGGAAACTTCTTGGTTGCTCAAAGCAAAG 854

Qy        781 gagatgggccctcccagttgggacagctgccatcgctcccatcattgctgctcaaggac 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 GAGATGGGCCCTCCCAGTTGGGACAGCTGCCATCCCATCATTCGTGTCGTCAAGGAC 914

Qy        841 gggaaaagcatcactcatgaagagaagaagaagatttttgggctgaaaagctgtgtactcctcca 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 GGGAAAAGCCATCACTCATGAAGAGAAGAGAGATTTTTGGGCTGAAAAGCTGTGTTACTCCTCCA 974

Qy        901 gatcctggttgctgttgtgttggtagaagtgtccagatgaaagcttcattcaaccccatc 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        975 GATCCTGGTTGCTGTCTGTTTGTTGGTAGAATGTCCAGATGAAAGCTGTCAATTCAACCCATC 1034

Qy        961 tgtgagaatgcaccttcaagagagtaccaagcagaagtccccgtgggcccttggtg 1020
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1035 TGTGAGAAGCCACCTTTCAGAGGTACCAAGGAAAGGCAGAATGCCCCGTGGGCTTGGTG 1094

Qy       1021 gttcacatggcccagacatctgcttgtggagaacagtgaaggcagaagagg 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1095 GTTCACATGGCCCCAGCATCTGCTTGTGGAGACAGCAGTGGAATGGGAGAGG 1154
```

```
Qy   1081 tttgggcctgacaccoagcactgttggtcctgaatgagaactgttgcctcagttcacaacctt 1140
          ||||||||||||||||||||||| ||| |||||||||||||| |||||||||||||||||||
Db   1155 TTTGGGCCTGACACCCAGCACTTGGTCTGAATGAGAACTGTGCCTCAGTTCACAACCTT   1214

Qy   1141 cgcagccacaagcagattcaaacccaagctcaacctcatccacccggaactcttcccctgctc 1200
          |||||||||||||  ||||||||||| |||||||||||||||||||||| |||||||||||||
Db   1215 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCTGCTC   1274

Qy   1201 accaagtttccgctgtaaggaaggaggggccccacccctcagtgtgttcacactggttcaggg 1260
          ||||| |||||||||||   ||||||||||||||||| ||||||||||||||||||||||||
Db   1275 ACCAGGTTTCCGCTGT---AAGGAGGGGGCCCCACCCTCAGTGTGTTCACACTGGTTCAGGG 1331

Qy   1321 tgcaatcctgaggaattcatagttgaggcgcttccaacttccagcagagccgtg        1380
          |||||||||||||||| || ||||||||| ||||||||||||||||||||||||||
Db   1332 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTTCCAACTTCCAGCAGAGCCGTG       1391

Qy   1381 caggagtacagaggaggagtgcgcaggaagaggggcccagcttccaacttccagcagtac    1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 CAGGAGTACAGAGGAGGAGTGCGCAGGAAGAGGGGCCCAGCTTCCAACTTCCAGCAGTAC    1451

Qy   1441 ccagaaatcatcttcctggaacaggggtctgccatccgatgaaagtgtcagt          1500
          ||||||| ||||||||||||||||||||| |||||||||||||||||||||||
Db   1452 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCGATGAAAGTGTCAGT         1511

Qy   1501 gccacacttgtcaacataagcccgacaacgtctctctactggactgttgtgaggcaca       1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512 GCCACACTTGTCAACATAAGCCCGACAACGTCTCTCTACTGGACTGTTGTGAGGCACA     1571

Qy   1561 tttggcaagtcgtgtgctcattacggagagacaggtgtccagggacaccctggct        1620
          ||||||||| |||||||||||||||||||||||| |||||||||||||||||||||
Db   1572 GCCAACACTTGTCAACATAAGCCCGACAACGTCTCTGTCTACTGGACTGGTTGAGGGCACA  1631

Qy   1621 gctgtgttgtctccacctgcacgcagcagatcaccacacacggggcttgccaagtatctcctgctg 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632 TTTTGGGCAGCTGTGCCGGTCATTACGGAGAACCAGGTTCCTGGGCACCCTGGCT       1691

Qy   1681 cagagagaacggcgccctggacatctttgggaaagccgcttcaaccctttgctgtggttgcc     1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGGCTTGCCAAGTATCTTGCTG  1751

Qy   1741 cccaacacagctcaaaggctggctccagcagtaccacaaccagtgccaggaggtcctgcac    1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752 CAGAGAGAACGGCGCCCTGGACATCTTTGGGAAAGCCGCTTCAACCCTTTGCTGGTGTTCC   1811

Qy   1801 cacatcagtatgattcctgccaaatgtcaggaaagggggctgaagtctccagtcctga       1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1812 CCCAACACAGCTCAAAGCCTGGCTCCAGCAGTACTACAACCAGTGCCAGGAGGTCCTGCAC  1871

Qy   1861 gtggaaagattgatcagttcgctgtcgcgacacatgtgattgatttggaaggttcagacctgt    1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872 CACATCAGTATGATTCCTGCCAAATGTCAGGAAAGGGGGCTGAGATCTCCAGTCCTGCA    1931

Qy   1921 ctggtgtcggcactgcaagcatgcgttggctgtgtgctgcgctgcgtcggtggtcggtgaa     1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932 GTGGAAAGATTGATCAGTTCGCTGTTCGCGAACATGTGATTTGAAGAGTTTCAGACCTGT    1991

Qy   1981 gtggtctattccgggacacatgccctgtccgaggcctctcggccgatggggaaagtgcc      2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1992 CTGGTCGTCGGCACTGCAAGCATGCGTTTGGCTGTGTGCTGCGCTGCGTCGGTGGTCGGTGAA 2051

Qy   2041 acctcctgatacatgaagccaacccctggaagatgttggaggaagcagtggaaag        2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2052 GTGGTCTATTCCGGGACACATGCCCTGTCCGAGGCCTCTCGGCCGATGGGGAAAGTGCC     2111

Qy   2101 acacacagacaaagtcccaagccaatcagctcgtggggatggaaccgagagttcatt       2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   2172 ACACACAGACAAAGTCCCAAGCCAATCAGCTCGTGGGGATGGAACCGAGAGTTCATT      2231

Qy   2161 atgctgaacacttcagccagcgctatgccaaggtccccctcttcagtcccaacttcagc      2220
          ||||||||||||||||||||||||||||||||||||||||| ||||| || |||||||||
Db   2232 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCAACTTCAGC    2291

Qy   2221 gaggaaagtgggagttgccttgaccacatgaaggtctgttggtcagttcccaacaatg      2280
          |||||||||||||||||||||||||||| |||||| ||||  |||||||||||||||||
Db   2292 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAAGGTCTGCCTTGGAGACTTTCCACAATG  2351

Qy   2281 cccaagctgattccccacctgaaagccctgtgtttgctgcggacatcgaaggaagatggggaag 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2352 CCCAAGCTGATTCCCCACCTGAAAGCCCTGTGTTTTGCTGCGGACATCGAAGGAAGATGGGGAAG 2411

Qy   2341 cgcgggagaaagcgggagctgcgcaggcgcgggcgcggcccctcctgtccaggagctggca      2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2412 CGCAGGGAGAAGCGGGAGCTGCGCAGGCGCGGGCGCGGCCCCTCCTGTCCAGGAGCTGGCA   2471

Qy   2401 ggcggcctggaggatggggggagcttcagccacagcagcagcgggcccacacagagaggaggc  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2472 GGCGGCCTGGAGGGAT-GGGGAGCCCTCAGCCACAGCAGCAGCGGGCCCACAGAGAGGAGGC   2530

Qy   2461 aagaaggtcagagcccagtga       2481
          |||||||||||||||||||||
Db   2531 AAGAAGGTCAGAGCCCAGTGA        2551
```

RESULT 6

LOCUS   AF308694   2893 bp   mRNA   PRI   27-FEB-2001
DEFINITION   Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION   AF308694
VERSION   AF308694.1 GI:10946488
KEYWORDS   .
SOURCE   gorilla.
  ORGANISM   Gorilla gorilla
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1 (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
             Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
             Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K.E., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
             Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
             Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED   11175785
REFERENCE   2 (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
             Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
             Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
             Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
             Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
             Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
             Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
             Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
             and Cannon-Albright,L.A.
  TITLE   Gorilla gorilla ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
             Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
             Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
             Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
             Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
             Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
             Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
             Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
             and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt

```
FEATURES          Location/Qualifiers
source            1..2893
                  /organism="Gorilla gorilla"
                  /db_xref="taxon:9593"
gene              1..2893
                  /gene="ELAC2"
CDS               1..2481
                  /gene="ELAC2"
                  /note="similar to Homo sapiens HPC2/ELAC2"
                  /codon_start=1
                  /product="ELAC2"
                  /protein_id="AAG24916.1"
                  /db_xref="GI:10946489"
                  /translation="MWALCSLLRSAAGRPMSQGRTISQAPARRERPRKDPLRHLRTRE
                  KRPGSGCSGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                  VRLDNIFLTRMHWSNVGGLSGMIILFKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                  IELAVRPHSAPEYEDETMTVYQIPIHSEQRGRHQPWQSPERPLSRLSPERSSDESSN
                  ENEPHLPHGVSQRGRVRDSSLSVVAFICKLHLKRGNFLVIKAKEMGLPVGTHAIAPIIA
                  AVKDGKSITHEGRETLAEELCTPDPGAARVVVECPDESPTIQPICENATFQRYQKAD
                  APVALVVHMAPESVLVDSRYQQWMERFGDTQHLVLNENCASVHNLRSHIQTQLNLI
                  HDIFPLITSFSPCKKEGPPLSVPMVQGECLLKYQLRPRREWQRDAIITQNPEEFIVEA
                  LQLPNFQQSVQEYRRSVQDVPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                  DTSLLDCGEGTFEQLCRHYGDQVDRVLGTLAAVFVSHLHADHHFGLLNLILLQREQAL
                  ASLGKPLHPLLLVVAPSQLKAWLQQYHHCQEVLFHISMIPAKCLGEGAEISSPAVERL
                  ISSLLRTCDLEEFQTCLVRNCKHAFGCALVHTSGNKKVVISGDTMPCEALVRMGKDATL
                  LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMINFSQRYAKVPLFSPNFN
                  EKVGVAFDHMKVCFGDFPYMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSGE
                  LAGGLEDGEPQQOKRAHTEEPQOAKVRAQ"
BASE COUNT        705 a        787 c        815 g        586 t
ORIGIN
```

```
Query Match       98.5%;  Best Local Similarity 99.0%;
Matches 2457;  Conservative 0;  Pred. No. 0;  Mismatches 24;  Indels 0;  Gaps 0;

                  Score 2442.6;  DB 89;  Length 2893;
```

```
Qy      1 atgtggcgcttgctctgtctgggtcgccgcggcc gaggagaagaccgagca                       60
          |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db      1 ATGTGGGGCCGCTTTGCTCCTCGCGTCCTGCGGTCTCGGCGGCACCATGTGGCAGGCAGC              60

Qy     61 accatatcgcaggcaccccgccgccgcaggagcccgctgcgggacctg                        120
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCATATCGCAGGCACCCGCCGCCGCCGCAGGAGACCCGCTCGGCACCTG                       120

Qy    121 cgcacgcgagagaagcgcgagacgcgaccgtcgggtgtgctccggaccg                      180
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCACGCGAGAGAAGCGCGAGACGCGTCGGGTGCTCCGGGACCGTGTACCTG                    180

Qy    181 caggtggttggcaagcgggtagccgggactccgggcgccgggcccaaaca                      240
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGGTTGGCAAGCGGGTAGCCGGGACTCCGGGCGCCGGGCCCAAACAC                      240

Qy    241 aaccggtatcttcttcaactgtggaagaggcttcagagaagacaacaagtta                    300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AACCGGTATCTTCTTCAACTGTGGAAGAGGCGTTCAGAGACTTCATGCAGGACACAGTTA            300

Qy    301 aaggttgctccgcctcgggacaacacatattcctgaccagcaatgcactggt                    360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGGTTGTTGGCCCTGGACAACACATATTCCTGACCAGGAATGCACTGGTTCTCAGTTA             360

Qy    361 ttaagtggaatgatctttacttttaaaggaaaacccgggcttcaaagtgtgtactttca             420
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTAAGTGGAATGATTCTTTACTTTTAAAGGAAAACCGGGCTTCCAAAGTGTGTACTTTCT           420

Qy    421 cctccacaactggaaaaatacctcgaagcaatacaaaataattttctggtgccatttgaaagga        480
          ||||| ||||| |||| |||||| ||| || ||||||||
Db    421 CCTCCACAGCTGGAAAAATACCTCGAAGCAATCAAAAATAATTTTCTGGTCCATTGAAAGGA         480

Qy    481 atagaaactggctgtcggcggcccacactctgccccaagaatacgaggatgaaaccatgacagtt       540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATAGAACTGGCTGTCGGCGGCCCCACACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT        540

Qy    541 taccagagatcccccatacacacagtgtgaaacaacaagaagggaggggagtcca                 600
          ||||||| ||| |||||||||||||||||||||||||||||||||||| |||||||
Db    541 TACCAGAGATCCCCATACACACAGTGTGAACACAGCAGGAGGAGGGGGAGTGGAAAGGCACCAGAGTCCA 600

Qy    601 gaaaggcctctcagcaggctcagtcctccagagcgagcagatcttcagagctcgagtcgaatgaaat    660
          ||||| |||||||||||||||| |||||||||||||||||||||| | | ||
Db    601 GAAAGGCCTCTCAGCAGGCTCAGTCCTCCAGAGCGAGCGATCTTCAGAGCTCGAGTGAAAT        660

Qy    661 gagccacacacacttccacactggttgttagccagagagagcaggtcaggggacttccttggtc       720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGCCACAGACACTTCCACACTGGTTGTTAGCCAGAGAGAGCAGGTCAGGGGACTCTTCCCTGGTC      720

Qy    721 gtagcttcattcgtgtaagcttcacttaaagagaaggaaacttcttggtgctcaaagtcaaag       780
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTAGCTTTCATTCGTGTAAGCTTCACTTAAAGAGAGGAAAACTTCTTGGTGCTCAAAGTCAAAG       780

Qy    781 gagatggggcctcccagttgggacagctgccatcgactccatcattgctgctgtcaaggac           840
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGATGGGGCCTCCCAGTTGGGACAGCTGCCATCGACTCCATCATTGCTGTCGTCAAGGAC           840

Qy    841 gggaaaagcatcactcatgaaggaagaggagatttggctgaaggagctgtgtactcctca           900
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGGAAAAGCATCACTCACTGAAGGAAGAGGATTTTGGGCTGAAGGAGCTGTGTACTCCTCCA        900

Qy    901 gatcctggttgctgcttttgtgtggtagaatgtccaagtgaagcttcattcaaaccatc             960
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GATCCTGGTTGCTGCTTTTGTGTGGTAGAATGTCCAGAGTGAAAGCTTCATTCAACCATC            960

Qy    961 tgtgagaatgccaacctttcaggaggttaccaagagaaagggggcccccgtggcctggtg          1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGTGAGAATGCCAACCTTTCAGGAGGTTACCAAGGAAAAGGCAGAATGCCCCGTGGCCTTGGTG      1020

Qy   1021 gttcacatggcccagcatctgcttgtgtggacacagcaggtaccagtggatgggagg            1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCCAGCATCTGCTTGTGTGGACACAGCAGGTACCAGTGGATGGGAGAG         1080

Qy   1081 tttgggcctgacacccagcacttggttcctgatgagaactgtgcctcagttcacaacctt        1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGGCCTGACACCCAGCACTTGGTTCCTGATGAGAACTGTGCCTCAGTTCACAACCTT       1140

Qy   1141 cgcagccacaagattcaaaaccagctcaaactcaatctcatccaccggacaactctccccgtc      1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAGCCACAAGATTCAAAACCAGCTCAAACTCAATCTCATCCACCGGACAACTTCCCCTGTC     1200

Qy   1201 accagttcctgtaggagctccgtgggcagagcccccacccctcagtgtgccatggtcagggtgaa    1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAGTTCCTGTAGGAGCTCCGTGGGCAGAGGGCCCCACCCTCAGTGTGCCATGGTCAGGGTGAA    1260

Qy   1261 tgcctctcaagtaccagctccgtcccaggagggaagggattgccattatact             1320
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGCCTCTCAAGTACCAGCTCCGTCCCAGGAGGGAATGGCCATGGTTCAGGGTGAA            1320

Qy   1321 tgcaatcctgaggaattcaatagttgaggcgctgcagctcccaactttccagagcgtg          1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGCCTCTCAAGTACCAGCTCCGTCCCAGGAGGGAATGGCCATTATCACT               1380

Qy   1381 caggagtacaaggagtgcgcaggacgcccagcccagccagagaaagaagtcagtac            1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGGAGTACAAGGAGTGCGCAGGACGTGCAGGACGTCCAGCCCAGAGAAAGAAGTCAGTAC     1440

Qy   1441 ccagaaatcatcttcctggaacagggtctgcacctccgatgaagattccgaaatgtcagt       1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGAAAATCATCTTCCTTGGAACAGGGTCTGCACATCCCCATGAAGATTTCGAAATGTCACT     1500

Qy   1501 gcacaacttgtcaacataagccccgacacgtctctgctactggactgtggagggcaca         1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCACAACTTGTCAACATTAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACG     1560

Qy   1561 ttgggcagctgtcggtcattacggacgaccagtgtgtgaccctggctggct           1620
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGGGGAGCTGTGCCGTCATTACGGACGACCAGTGTTGGCACGGTCCTGGCACCCTGGCT     1620

Qy   1621 gctgtgttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg
          ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1621 GCTGGTGTTGGTGTCCCACCTGCACGCAGCAGATCACCACAGGGGCTTGCTAAATATCTTGCTG 1680
          |||||||||||||| |||||||                       ||  ||||||||||||||
Qy   1681 cagagagaaacggccctggcatctttgggaggaccgcttcaccttgctgtgggttgcc 1740
Db   1681 CAGAGAGAACAAGCCTTGGCACATCTTTTGGGAAAGCCCCTTCACCCTTTGCGTGGTTGCC 1740
          |||||| |||        ||||||||||                   ||||||||||||||
Qy   1741 cccaaccagctcaaagcctggctccagcgagtaccacacagtgccagaggagtcctgcac 1800
Db   1741 CCCAGGCCAGCTCAAAAGCCTGGCTCCAGCGAGTACCACACCAGTGCCTCGAC 1800
          |||||||||||||||||||||||||||         |||||||||||||||
Qy   1801 cacatcagtatgtccggggacacatgcctcaggaaggtgagatctccagtcctgca 1860
Db   1801 CACATCAGTATGATCCTTGCCAAATGCCTCAGGAAGGGCTCGAGATCTCCTCCA 1860
          |||||||||| |||||||||||||||||||||||        |||||||||||
Qy   1861 gtggaaagattgatcagttgctgcgggaaagacacttcagaactgt 1920
Db   1861 GTGGAAAGATTGATCAGTTCGCTGCGGGAAATGCCTTCAGGAGAGTCATCCTCGA 1920
          ||||||||||||||||||||||||||||                 |||||||||
Qy   1921 ctggtgcggcactgcaagcatgcggtttcgcctgtgcacacctctggctgaa 1980
Db   1921 CTGGTGCGGCACTGCAACCATGGTTTCGCTGTGTGGTGTTGCACACCTCTGCCTGAAA 1980
          |||||||||||||||||||||||||||||||                ||||||||
Qy   1981 gtggtctattccgggacaccaggtgcgtatgccaagtggggaaagatgcc 2040
Db   1981 GTGGTCTATTCCGGGGACACCAGGTGCCTTCGCGACGTCCGCCATGGGGAAAGATGCC 2040
          |||||||||||||| |||||||||||||||||               |||||||||
Qy   2041 accctcctgataccatgagcagcccctggaagctcgtcctcccgctcgg 2100
Db   2041 ACCCTCCTGATACATGAAGCCACCCTTGGAAGATCGTTTGGAAGACAGTGGGAAAAG 2100
          |||||||||||   ||||||||||||||||||||                |||||
Qy   2101 acacacagcacaacgtcccaagccaagccatcagcgttgggggatgaacgcggagttcatt 2160
Db   2101 ACACACACAACGTCCCAACCATCGGTTTGCTTCAGCGTGGGATGAACGCGGAGTTCATT 2160
          |||||||||||||||||||                    |||||||||||||||||||
Qy   2161 atgctgaccacttcaggacaccgtatgcgcaggtgtccccccaacttcagc 2220
Db   2161 ATGCTGAACCACTTCAGACACCGTATCGCCAAGTCCCCTCGTCTTCTTCAGCCAACTTCAAC 2220
          ||||||| ||||||||||||||||| |||||||                 |||||||| |
Qy   2221 gagaaagtgggaggagttgcctttgaccaacaatgaagcgttaccacaatg 2280
Db   2221 GAGAAAGTGGGAGAGTTGCCTTTGACCAACGCTATGAAGGTCTCGCTTGCCCAACAATG 2280
          |||||||||||| |||||||||||||||||               |||||||||||
Qy   2281 cccaagctgattcccccactgaaagccctgttctgtcgggcaacatcgaaggagagtggaaggag 2340
Db   2281 CCCAAGCTGATTCCCCACTGAAAAGCCCTGTTTGCCGGCGACAATCGAGAGGATGGAGGAG 2340
          |||||||||||||||| |||||||||||                 ||||||||||||||
Qy   2341 cgcaggaggagaggacgcggagctgcggccggcgcggccgccctcctgtccaggggagctggca 2400
Db   2341 CGCAGGAGGAAGCGGGACCTGCGGCAGTGCGGGCGGCCGCCTCTTCTGTCTCCGGGGAGCTGGCA 2400
          |||||||||  ||| ||||||||||||||||| ||                 |||||||||
Qy   2401 gcgggcctgaggggatgggaggaggaacgcgggcgggccccaacagaggagccacaaggcc 2460
Db   2401 GCGGGCCTGGAGGATGGGGAGGGAGGAGCTGCGGATGGAGCTCAGGAGAACGGGCCACAGGCC 2460
          ||||||||||||||| |||||  ||||||||||                ||||||||||||
Qy   2461 aagaaggtcaagagcccagtga 2481
Db   2461 AAGAAGGTCAGAGCCCAGTGA 2481
          |||||||||| ||||||||||
```

RESULT 7

```
LOCUS       AF308696    2712 bp    mRNA       ROD    27-FEB-2001
DEFINITION  Mus musculus ELAC2 mRNA, complete cds.
ACCESSION   AF308696
VERSION     AF308696.2  GI:11992378
KEYWORDS    
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
```

Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED   11175785
REFERENCE   2 (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
  TITLE   Mouse ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
REFERENCE   4 (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-DEC-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
  REMARK   Sequence update by submitter
COMMENT   On Dec 26, 2000 this sequence version replaced gi:10946492.
FEATURES   Location/Qualifiers
     source   1..2712
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="fetus"
                     /note="similar to Homo sapiens HPC2/ELAC2"
     CDS   15..2510
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24918.2"
                     /db_xref="GI:11992379"
                     /translation="MWALRSLLRPGLGLRFIMSQGSARRPPSKDPLRHLRTEKRGPGP
                     GGPNVILQVVAAGGRDAGAALIYPESERNYRFLPNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLRKCVLSGPPQLEKYLEAIKIFSFGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERCGKQOPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVFCKLHLRGNFLVLKAFELGPVGTAAIAPIIAAVKD
                     GKSTIYEGRREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDFKRYQAEADAPVA
                     LVHHAPESVLIDSRYQOWMERFGPDTOHLILNENCPSVHNLRSHKIQIOLSIIHPDI
                     FPQLISFYSKEEGSLLSVPTVRGECLLKYQLPRREWQRDTTPDCNTDEFIAEALEP
                     SFQESVEEYRKNVEONPAEAEKRSOYPEIVFLGTCSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQOIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTOLRAWLQOYHNCQEILHHVSMIPAKCLQKGAEVSNTLERLISLL"

```
BASE COUNT      659 a    741 c    748 g    564 t
ORIGIN

Query Match     68.0%;   Score 1686.2;  DB 94;  Length 2712;
Best Local Similarity   82.2%;   Pred. No. 0;  Matches 408;
Matches 1982;  Conservative    0;  Mismatches    408;  Indels   21;  Gaps    3;

Db                  LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLHE
                    ATLEDGLEEEAVEKTHSTTSQAINVGMRNAGFIMLNHFSQRYAKIPLFSPDFNEKVG
                    IAFDIHMKVCFGDFPVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
Qy                  EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"

Qy    58  cgcacacatatcgcaggcgaccacccgcccgcccgcccgagcgccgcggcgcgcac 117
          |||||||||||| |||| | |||||||||||||||| ||||||| |||| ||| |||
Db    54  CGCACCAATGTGCAGGGTTCGGCTCGTCGGCCCATCCAAAGACCCACTGCGACAC 113

Qy   118  ctgcgcacgcggagagaagcgcggacgtcgggtgtgctccggcgcggcccaaaacc 177
          ||||| ||| |||| ||| | || | ||||| | |||||  |||||| ||||||  |
Db   114  CTGGCTACGCGGAGAAGCGCGGGCC------GGGTCCCGGGGCCCGAACACCGTGTAC 167

Qy   178  ctgcaggtggtcagcggccggggggactcgggccgctctacgtctctccgag 237
          ||||||||||||||||| | ||| ||| |  ||| | | ||| ||||| ||||
Db   168  CTGCAGGTGGTGGGCGGCGGGGGGGACCCCGGGGGCCTGCTCTCTATGTCTTCTGGAA 227

Qy   238  ttcaaccgtatctcttcaactgtggaaaaaatacctcgaagagaagcacaag 297
          |||||||||| |  ||||  |  ||| |||||||| | |||| ||| ||||||
Db   228  TACAACAGGTACCTTTTAACTGCGAGAAGCGGTCCAACGACTTATGCAGGAAACACAAA 287

Qy   298  ttaaaggttgctccgtggacaacatatttcctgacacgtgtctaatgttggg 357
          |||||||||||||| ||||||||| |||||| ||||||||||| |||| |||
Db   288  CTGAAAGTCCTGGCTTGGACAACATCTTTCCTGACTCGGATGCCATTGGTCATTGGG 347

Qy   358  ggcttaagtcggaatgattcttacttttaaaaggaaaccggcttccaaagtgtactttct 417
          |||||| |||||| |||||||||||| |||||||||||||| ||||||| |||| ||| ||
Db   348  GGGTTGTGTGGAATGANTTTAACTTTAAAAGGAAACCGGGCTTCCCAAATGTGGTTCTGTCT 407

Qy   418  ggacctccacaactggaaaaaatacctcgaagcaatcaaaaatatttctgtgtccattgaa 477
          |||||| ||| | |||||||||||| |  || |  |||||| ||||||| |||| ||||||
Db   408  GGACCACCACAGCTGGAGAAATATCTAGAAGCAATCAAAAATATTTTCTGGTGTCAATTGAAA 467

Qy   478  ggaatagaactggcgtgtcgcgcccccagaaagggatgaaaccatgaca 537
          |||||| ||||| ||||| | ||||| ||||| ||| ||||||||| ||
Db   468  GGAATAGAACTGGCGCCGTGCCGGCCTCACTCTGCACCAGGAATACAAGGAITGAGGACCATGACT 527

Qy   538  gtttaccagatccccatacacagtgaacacacccaaccatggcaagt 597
          ||||| ||||||||| |  ||||| || || || ||||||||||| |
Db   528  GTTTACCAGGTCCCTANCCACAGTGAACGGAGGTGGAAAGCAACAGCCATCCCAGAGC 587

Qy   598  ccagaaaggcctctcagaggagctcagtccagagtctccagtcgaatgaa 657
          || ||||||||||| ||| ||||||| ||||||| |||| ||| | |||
Db   588  CCCAGAACACTCTCCAACAGGCTCAGTCCAAACACTCTGGAICCAGCTGAA 647

Qy   658  aatgagccacacacctccaca------------tggtgttagcgcagga 705
          |||| ||| || || |||| |          || |  ||||| | |||
Db   648  AATGGGCAGTGCGCTCCCTCCAGAGACGTGCCCACAGTAAICAGAITGGGGGCAGG 707

Qy   706  gactcttccctggtcgtagcttcatcgtgaagcttaaagaggaggaaacttcttg 765
          || ||||||| || || |  || || | || || || ||||||||||||||||||
Db   708  GACCCCTCCTTAGTGTGAGCTTTGTCGCAAGCTTCACTGAGGAAAGGAAAACTCTTG 767

Qy   766  gtgctcaaagcaaggagagtgggcctcccagcctgccatcctcccatcatt 825
          | |||||||||||||||||||| |||| || ||||||||| || |||| ||
Db   768  GTGCTTAAAGCAAAGGGAGCTGGGCCTTCCTGTTGGGACACICCGCCACCCATCATT 827

Qy   826  gctgctgtcaaggacggagaaagcatcacactcaatggaagatttggctggagcgt 885
          ||||||||||||| | ||| |||| | || ||||||||||||| |||||||| |||
Db   828  GCTGCTGTTCAAGGACGGGAAGAGTATACACTCATGGAAGGAGATTGCTCTGCAGAG 887

Qy   886  ctgtgttactcctccagaatcctggttctgtctgctttgttggttagaatgtccagatgaaagc 945
          |||||  ||||| ||| |  | |  | ||||| | || | ||||  || |  ||||  |||||||
Db   888  CTTTGTACACCCCCAGAITCCTGGTCTCTGTCGTCTTACTCTGGGGGATAACGCGTTCTACTCGG 947

Qy   946  ttcattcaaccatctgtgaggaatgccacccttcagaggtaccaaggagaagagagatgcc 1005
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   948  TTCATCCATCCTGCCCATCTGTGAICGACGACACCTTTPAAAAGGTACCAGGGGCTGACATCGA 1007

Qy  1006  ccggcgcttggtgttcacatggccccagcatctgtgcttgttggacgcaggtacgcttc 1065
          |||| |  ||| |  |||| ||  || || |||| ||  ||||| | |||||||||||| |
Db  1008  CCTGTGGGCCTGGTGTGGCTCGGGCCCTGACACACACACTGATGATGACAGCAGATACAGAG 1067

Qy  1066  cagtggatggagaggtttggggctgacaccaccagcaactggaggtcctgaatgaaactg 1125
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1068  CAGTGGAITGGAGAGGTTCGGGGCCCTGACACCACCACCAGTATCCAACAGCCC 1127

Qy  1126  tcagttcacacaacctccgcagcacacaagctcaactggtctgttctggacagtggaccac 1185
          |||||||| |||| |||| ||||| |||| | | ||||||||||||||||||||||| || |
Db  1128  TCGGTCCACAAACCTGCGCAIGCCAACAGAITTCAGCACTCCAGCCCTCATCCACCCTGAC 1187

Qy  1186  atcttccccctgctccacacagttccgctgtaagaagagagggcccacacctcagtgtcc 1245
          |||||||| |||||||| |||||||| | || | || || || |||| | || ||| |||
Db  1188  ATCTTCCCCCAIGCTTACCAIGCITCTTATACTAIGAGGAIGAAAGGTCACCCICCAGCGTGCCA 1247

Qy  1246  atggttcaggtgaatgctccaagtaccagctccgtcccaggtacccggtgaatgcc 1305
          |||||||||||||||||||| || |  || | |||| || |||| | || ||| ||
Db  1248  ACAGTTCGGGTGAATGCTCCTCAAAGITCATCAGCTCCCCCAAAGAGGTCACCTCCAGAG 1307

Qy  1306  gatgccattattacttgcaatcctgaggagaattcataagttagagaagcttcccaac 1365
          ||||||||| | || |||||||||||||||| || | || |||| || || |||||| |
Db  1308  GATACCACCACCCGACTGCAAITACTGACATAICTGAAGAITTCATTAGCTAGGGACTTTGGAGCTCCCCAGT 1367

Qy  1366  ttccagcagacgtgcaggagtacaggagagtggcgcagagacaggtgagcttccaac 1425
          ||||||||| |||||||||||||||||||||| |  || | ||||||||||||||||
Db  1368  TTCCAGGACGTGTTGGAGGAITACGGGAGAGTGGCGAAAAACCCAGGCAGCTCCCCACAGAG 1427

Qy  1426  aaaagaaagtcagtacccagaaaaatcatctcctttggaaacgttgcgctgtcattccg 1485
          |||||||||||| | | ||||||| ||| ||||||||||| | | | |||||| ||| |
Db  1428  AAAAGAAAGCCACGATATCCTGAAAITTGTCTTCCCGGGGTCGTGCCGGTCIATCCAATGAAG 1487

Qy  1486  attcgaaatgtcagtgccacacttgtcacataagcccgacacagtctcctactggac 1545
          |||||||||||||||||||||||||| | |||| ||||||||| | ||| || ||||
Db  1488  ATCGAAATGTCAGITCCACACTGTCACAACCTAGGCGACAGGGICAGGTCIGCTGGAT 1547

Qy  1546  tgtggtggagggcacattgggcagcgtgtgcgctcattacggagaccaaggtggactc 1605
          |||| |||||||||||||||| || ||||| | || ||||| ||||||||||  ||||
Db  1548  TGTGGAGGAAGGCAGCACTTTGGGGCAGTTGTGTCCGTCCAITTACGGAITCAACAGCTC 1607

Qy  1606  ctgggcacccctggctgctgtgttgtttgtccacctgcagacaacacagggcttg 1665
          ||||||||||| || | ||||| ||||| || |  |  |||||||| || || ||
Db  1608  TTATGCAGGCCTCACCGGCTGTTTGTGTCCCCACCTGCAIGAAATGCCTTCAGAGGGGCTTG 1667

Qy  1666  ccaagtgtatcttgctgacagagagaacgcgccttggcatctttgggaaagcgcttcaccct 1725
          ||||||| || |  || | ||||||| || || ||| |  || |||| |||||| ||||||
Db  1668  CTGAAITATCTTGITTGACAITACTTACTGGCAICATCTCTCGGGGAAAACCCTTCCAGCCC 1727

Qy  1726  ttgctggttggttgccccacagcaaagcctggctcacagcaagtaccacacaagtgc 1785
          ||||| ||| ||  ||||| ||| |||| |  ||| ||| | | | |||| |  |||
Db  1728  TTGCTTGTGTGGGTCCCCACAITGITAAITCTGGGTCCGCCTATCANCAACACITGC 1787

Qy  1786  ctgggcacccctggctgctgtcatctgatgattccccgaccgaagtatcacacggcag 1845
          || ||||| || || ||||||| |   || | | ||||||  | | | || |||| ||
Db  1788  CAGGAITCTTGCGAGGCTCAGITGAGTITGGAGITTCCTCGAATGCCTTCCAGAGGCAGAG 1847

Qy  1846  atctccagtcctgcagtggaaagattgatcagttcgctgttcgtaacatgtgattgaa 1905
          |||||||||||| || |||||| | | | ||||| ||||  || ||||| ||| |||||
Db  1848  GTCTCCAAITACTACAITTGAITAAITAGCTTGCTTGGCTTGGAAAICAITGTGACTTTAGAA 1907

Qy  1906  gagttcagagcctgtctgtggtcgggcactcaagcatgcgtttggctggtgcctggtgcac 1965
          ||||| ||| || || |||| | |||  || |||| ||| ||| |||| || | ||| | |
Db  1908  GAATTTCAGAGCCTGITTGGGTCCTGTGITACCGGCACTCGCAAGCATGGCTTTTGGCTGGACGTACAT 1967

Qy  1966  acctctggctggaagtggtctattccgggggacaccatgcctctgaggcggct 2025
          |||||||||||||||||||| | |||| |||||| || |||||||| || |||
Db  1968  TCAITCIGGCTGGAAAGTCGITCTCTACTCGGGGAIIACCATGCGCTCTGTGAGCTCTTGGTTCCAG 2027

Qy  2026  atggggaaagatgccacccttcctgaatacatgaagggaatggtttggaagag 2085
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2028  ...
```

```
Db    2028  ATGGGGAAAGATCCACCCTCCTTGATACATGAAGCCACTCTGAGGATGGCTTGGAAGAG
                 ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    2086  gaagcagtggaaagacaacacagcacaacgtcccaagcatcagcgtgggatgcggatg    2087
Db    2088  GAAGCAGTAGAGAAGAGACACACACACCACCAAGCCTCCCAGGCCATTAATGTGGGATGCGGAGG    2147
Qy    2146  aacgcggagttcattatgctgaccaactttcagcagccgctatgccaagtcccctctc    2205
Db    2148  AATGCGGGGTTCATACATGCTGAACCACTTCAGTCAGCGGTACGCCCAAGATCCCCTTTTC    2207
Qy    2266  gactttccaacaatgcccaagctgattccccccactgaaagccctggtgctggcgacatc    2325
Db    2268  GACTTCCGACAGTGCCCAAAGCTGATTCCCACTGAAGGCTCCGTTTGCAGGTGACATT    2327
Qy    2206  agccccaacttcagcgagaaagtgggagttgccttgacacatgaaggtctgctgttgga    2265
Db    2208  AGCCCTGACTTCAACGAGAAAGTTGGCANCGCCTTTGACCANTGAAAGGTCTGCTTTTGGA    2267
Qy    2326  gaggagatggtggaagaaggagggggaaggagcgtgcggcccgcgcggccgcctcctg    2385
Db    2328  GAAGAGATGGTGGAAGAAGGAGCTACGGCTGGTTGCGAGCAGCGCCTCCTG    2387
Qy    2446  gaggagcccaca    2456
Db    2445  GATGAACCACA    2455
Qy    2386  tccaggggagctggcaggcgccctgagagagatggggaagggcggcaccaca    2445
Db    2388  ACC---CAGCAGCGAGCACCCAGAGAGACAGAGAACCCCACAGAGAACCGGGCCCACACA    2444

RESULT    8
LOCUS        AC005277    118788 bp    DNA    PRI    23-JUL-1998
DEFINITION   Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
ACCESSION    AC005277
VERSION      AC005277.1  GI:3337311
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 118788)
  AUTHORS    Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 17, clone hRPK.597_M_12
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 118788)
  AUTHORS    Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
             Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
             Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
             Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
             FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
             Gensheimer,S., Geralgery,K., Gilmartin,T., Grant,G., Hagos,B.,
             Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
             Macdonald,P., Marquis,N., Marris,W., McGurk,A., McKernan,K.,
             Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
             Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
             Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
             Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
             Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
             Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
             Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    3  (bases 1 to 118788)
  AUTHORS    Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
             Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
             Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
             Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
             FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
             Gensheimer,S., Geralgery,K., Gilmartin,T., Grant,G., Hagos,B.,
             Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
```

```
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
             Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
             Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
             Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
             Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
             Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
             Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Jul 23, 1998 this sequence version replaced gi:3335015.
             All repeats were identified using RepeatMasker; Smit, A.F.A. &
             Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html

             Only the first 118.8 kilobases of this clone are being submitted.
             The remainder overlaps accession number AC005274 (WICGR project
             L350).
FEATURES             Location/Qualifiers
     source          1..118788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.597_M_12"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   795..884
                     /rpt_family="MIR"
     repeat_region   complement(932..1000)
                     /rpt_family="MIR"
     repeat_region   2044..2076
                     /rpt_family="AT_rich"
     repeat_region   2237..2713
                     /rpt_family="MER89"
     repeat_region   complement(4282..4321)
                     /rpt_family="(TAAA)n"
     repeat_region   2718..3477
                     /rpt_family="L2"
     repeat_region   complement(3478..3821)
                     /rpt_family="MLT1A1"
     repeat_region   3822..4007
                     /rpt_family="L2"
     repeat_region   complement(4015..4281)
                     /rpt_family="MER89"
     repeat_region   complement(4328..4608)
                     /rpt_family="AluSx"
     repeat_region   complement(4609..4895)
                     /rpt_family="MER89"
     repeat_region   complement(4896..5186)
                     /rpt_family="AluSp"
     repeat_region   5219..5479
                     /rpt_family="L2"
     repeat_region   complement(5481..5740)
                     /rpt_family="MER34"
     repeat_region   complement(5963..6275)
                     /rpt_family="AluSp"
     repeat_region   complement(6411..6494)
                     /rpt_family="MER5A"
     repeat_region   6830..7013
                     /rpt_family="MLT1C"
     repeat_region   7014..7318
                     /rpt_family="AluYb8"
     repeat_region   7319..7581
                     /rpt_family="MLT1C"
     repeat_region   7721..8032
                     /rpt_family="AluJb"
     repeat_region   8039..8333
                     /rpt_family="AluSq"
     repeat_region   9431..9460
                     /rpt_family="AT_rich"
     repeat_region   9662..9829
                     /rpt_family="MIR"
```

```
repeat_region       complement(11302..11728)
repeat_region       /rpt_family="MER39"
repeat_region       complement(11824..11884)
repeat_region       /rpt_family="LTR16C"
repeat_region       complement(11928..12108)
repeat_region       /rpt_family="LTR16C"
repeat_region       12453..12725
repeat_region       /rpt_family="AluJb"
repeat_region       13377..13785
repeat_region       /rpt_family="MSTB"
repeat_region       14118..14153
repeat_region       /rpt_family="(CA)n"
repeat_region       complement(15320..15620)
repeat_region       /rpt_family="AluJo"
repeat_region       15683..15759
repeat_region       /rpt_family="(CA)n"
repeat_region       complement(15838..15893)
repeat_region       /rpt_family="MER91A"
repeat_region       complement(15915..16080)
repeat_region       /rpt_family="MER91A"
repeat_region       complement(16474..16544)
repeat_region       /rpt_family="L2"
repeat_region       17907..17920
repeat_region       /rpt_family="L2"
repeat_region       18271..18311
repeat_region       /rpt_family="L1ME"
repeat_region       complement(18358..18529)
repeat_region       /rpt_family="L2"
repeat_region       18530..18560
repeat_region       /rpt_family="AluJb"
repeat_region       complement(18561..18681)
repeat_region       /rpt_family="(CA)n"
repeat_region       18760..19065
repeat_region       /rpt_family="AluJb"
repeat_region       complement(19638..19896)
repeat_region       /rpt_family="AluSg"
repeat_region       complement(19971..20260)
repeat_region       /rpt_family="MIR"
repeat_region       2063..20552
repeat_region       /rpt_family="L1ME3A"
repeat_region       complement(20553..20816)
repeat_region       /rpt_family="MER58B"
repeat_region       complement(20821..21069)
repeat_region       /rpt_family="L1ME3A"
repeat_region       complement(21071..21100)
repeat_region       /rpt_family="L1M4"
repeat_region       complement(21101..21338)
repeat_region       /rpt_family="(TAAA)n"
repeat_region       complement(21348..21598)
repeat_region       /rpt_family="AluSx"
repeat_region       21799..22104
repeat_region       /rpt_family="L1M4"
repeat_region       22151..22562
repeat_region       /rpt_family="AluJo"
repeat_region       23978..24287
repeat_region       /rpt_family="MSTA"
repeat_region       24852..25199
repeat_region       /rpt_family="L1M4"
repeat_region       complement(25697..25985)
repeat_region       /rpt_family="AluSq"
repeat_region       25816..25895
repeat_region       /rpt_family="L2"
unsure              25816..25895
                    /rpt_family="MLT1D"
                    /note="Single-stranded coverage."
repeat_region       complement(25994..26099)
repeat_region       /rpt_family="MLT1D"
repeat_region       complement(26122..26291)
repeat_region       /rpt_family="purine-rich"
repeat_region       26312..26637
repeat_region       /rpt_family="(CAAAA)n"
repeat_region       26456..26480
repeat_region       /rpt_family="(GGGAA)n"
repeat_region       complement(26576..26781)
repeat_region       /rpt_family="MER6B"
repeat_region       complement(26912..26990)
```

```
repeat_region       26991..27290
repeat_region       /rpt_family="AluSx"
repeat_region       complement(27291..27975)
repeat_region       /rpt_family="MER21B"
repeat_region       28034..28197
repeat_region       /rpt_family="LTR28"
repeat_region       28266..28795
repeat_region       /rpt_family="LTR28"
repeat_region       28843..28916
repeat_region       /rpt_family="LTR28"
repeat_region       28920..29005
repeat_region       /rpt_family="LTR28"
repeat_region       29140..29233
repeat_region       /rpt_family="MIR"
```

```
                    /rpt_family="MER21B"
```

Query Match   10.0%;   Score 247.4;   DB 85;   Length 118788;
Best Local Similarity   97.7%;   Pred. No. 3.9e-48;
Matches 251; Conservative   0; Mismatches   6;   Indels   0;   Gaps   0;

Qy   1 atgtgggcgcttgctcggctgctgggcgtcccggcggccggaagcaccatgtccgcaggggagc   60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78562 ATGTGGGCGCGCTTTGCTCGGCTGCTGGGCGTCCCGGCGGCCGGAAGCACCATGTCCGCAGGGAGGAGC 78621

Qy   61 accatatcgcaaggcgaccccgccgccgcgccgagacgcgccgaaggacccgtgcggcacctg 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78622 ACCATATCGCAGGCGACCACCCGCCGCCGCGCCGAGCGCGCCGCCAAGGACCCGTGCGGCACCTG 78681

Qy   121 cgcacgcgagagaagcgcggaccgtcgggggtgctccggcggcgcccaaacacacgtgtacctg 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78682 CGCACGCGAGAGAAGCGCGGACCGTCGGGGGTGCTCCGGCGGCGCCCAAAACACACGTGTACCTG 78741

Qy   181 caggtggtggcaacgggttagccgggacttcgggggcgcgcgctcttacgtcttctctcggagttc 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78742 CAGGTGGTGGCAACGGGTTAGCCGGGGGTAGCCCGGGACTCGGGCGCGCGCTCTACGTCTTCTCCGAGTTC 78801

Qy   241 aaccggtatcttctttcaa 257
            ||||||||||| ||| |||
Db   78802 AACCGGTCAGTCAACGA 78818

REFERENCE
  AUTHORS   Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.,
            and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES   Location/Qualifiers
     source   1..740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     mRNA   301..>439
                     /number=17
     exon   301..>439
                     /product="putative prostate cancer susceptibility protein"
                     HPC2/ELAC2
     CDS   <301..>439
                     /codon_start=2
                     /product="putative prostate cancer susceptibility protein"
                     HPC2/ELAC2
                     /protein_id="AAG24440.1"
                     /db_xref="GI:10800931"
                     /translation="PDYSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADH
                     HT"
BASE COUNT   154 a   188 c   173 g   225 t
ORIGIN

Query Match   5.9%; Score 146.2; DB 89; Length 740;
Best Local Similarity   89.7%; Pred. No. 5.7e-24;
Matches 157; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy   1486 attcgaaatgtcaagtgccacacttgtcaacataagcccccgaacacgtgtctctgctactggac 1545
          |||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||| |||
Db    266 AGTGGCAGTGACTCTCTCTCTCTTCCTCGAGCCCGGACAGCGTCTTCTGCTTACCTGGAC 325

Qy   1546 tgtggtgagggcacacattgggcagctgtgccgtgtcattacggagaccaagggtc 1605
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    326 TGTGGTGAGGGCACATTGGGCAGCTGTGCCGTGTCATTACGGAGACCAAGGGTC 385

Qy   1606 ctgggcaccctggctgtgtgtttgttctccccacctgcacgcagatcaccacacgg 1660
          |||||||||||||||| |||||| |||||||||||||||||||||||||||||||||
Db    386 CTGGGCACCCTGGCTGTGCTGTTTGTGTCCCACCTGCACGCAGATCACCACACGG 440
```

RESULT 10
LOCUS   AF348157   34593 bp   DNA   ROD   04-APR-2001
DEFINITION Mus musculus putative prostate cancer susceptibility protein
            (Elac2) gene, complete cds, alternatively spliced.
ACCESSION   AF348157
VERSION   AF348157.1   GI:13540341
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   Mouse Elac2-containing genomic DNA
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 34593)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES   Location/Qualifiers
     source   1..34593
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene   <8470..>31277
                     /gene="Elac2"
     mRNA   join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27930,28039..28553,28652..>30639)
                     /gene="Elac2"
                     /note="alternatively spliced 24 exon form"
                     /product="putative prostate cancer susceptibility protein"
     mRNA   join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27930,28039..28553,29101..29179,29710..
                     31242..>31277)
                     /gene="Elac2"
                     /note="alternatively spliced 25 exon form"
                     /product="putative prostate cancer susceptibility protein"
     CDS   join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27930,28039..28553,28652..30639)
                     /gene="Elac2"
                     /codon_start=1
                     /note="24 exon form"
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAK29421.1"
                     /db_xref="GI:13540343"
                     /translation="MWALRSLLRPLGLRFIMSQGSARRPRPPKDPLRLHRTREKRGPGP
                     GGRNTVYLQVVAAGGRDAGAALYVFSERVNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLRKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQOPSQSPRTSPNRLSPKQSSDSGSAAENGCQPPE
                     DSSAGANRRAWGRDPSLVVAFVCKLHIRGNFLVLKAKELGLPVGTAAAAVAKD
                     GKSIYEGRRIAAEELCTPDDPGLVFIVVECPDDGFILPICENDFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQOWMERFGPDTOHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTFSFYSKEEGSTLSVPTVRGECLLKTQOLRPKREWORDTTLDCNTDFIAERALELP
                     SFQESVEEVRKNVOENPAPAEKRSQYPEIVFLGTCSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCCGEGTFGQLCRHYGQOJDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQOYHNICOEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFOTCLVRHCKHAFGCALVHSSGWKVVYSGDPMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKRTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFGDPTVPKLIFIPPLKALFACDIEEMVERREKRELRLVRAALLTQOASP
                     EDREPQOKRAHTDEDPHSEPQOSKKVRAQ"
     CDS   join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27930,28039..28553,28853..29101,
                     31242..>31277)
                     /gene="Elac2"
                     /codon_start=1
                     /note="25 exon form"
                     /product="putative prostate cancer susceptibility protein"

Left column

```
                     /protein_id="AAK29420.1"
                     /db_xref="GI:13540342"
                     /translation="MWALRSLLRPLGLRTMSOGSARRPRPKDPLRHLRTRKRGPGP
                     GGPNYVLQVVAAGRDAGAALYVFSEYNRYLFNCGEGVQORLMQEHKLKVARLDNIFL
                     TRMHWSNVGLCGMILTLKETGIKPKCVLSGPPQOLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMYVQVPIHSERRCGKQOPSQSPRTSPNRLSPKQSSDSGSSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKIHLRKGNFLVLKAKELGLPVGTPAAIAPIIAAVKD
                     GKSITYEGRETAAEELCTPPDGLVFIVVEGCDEGFILPICENDTFKRYQAEADAPVA
                     LVHIAPESVLIDSRYQQWMERFGPDTQHLLINENCPSVHNLRSHKIQTOLSLIHPDI
                     FPQLTSFYSKEEGGSTLSVPPVRGEECLLKYQLRPKREWQRDTTLDCNTDEFIAEALEP
                     SFQESVEEYRRNVQENPAPAEKRSQPFEIVEGTLGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSITAVEFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQLLVVAPTOLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFQTCLVRICKHAFGCALVHSSGMKVVIYSGDTMPCEAIVQDMGKIDATLLHE
                     ATLEDGLEEAAVEKTHSTTSQAINVGMRMNAEFIMINHFSQRYAKIPLFSPDENEKVG
                     IAFDHMKVCGDFPTVPKLIPPLKALFAGDIEEMVEREREKRELRIVRAALLTQQADSP
                     EDREFOQKRAHTDEPHSPQOSKKESVANTLGARV"

BASE COUNT     8641 a   8030 c   8142 g   9780 t
ORIGIN

Query Match     79.1%;  Score 144.4;  DB 94;  Length 34593;
Best Local Similarity   79.1%;  Pred. No. 1.1e-23;
Matches  201;  Conservative   0;  Mismatches  41;  Indels  12;  Gaps  2;

Qy   2237 ccttgaccacatgaaggtctgctcttggagacttccaacaatgcccaagctgattcccc 2296
          |||   |||||| |||||||| ||| ||||||||||||| ||||||||||||| ||||||
Db  30389 CCTGTCTGTCCTTCCAGGTCTCGCTTGGAGACTTCCGACAGTGCCCAAGCTGATTCCCC 30448

Qy   2297 cactgaaagccctgttgctggcgacatcgaggagatgagggaggaaggcggg 2356
          ||||||| ||||||||| |||| ||||| ||||| ||| ||||||| ||| ||
Db  30449 CACTGAAGGCCCTGTTGCCAGGTGACAATTGGTGGAACGGCCAGGAGGAAGAGG 30508

Qy   2357 agctgggcggcaggtgcgggcggcgcccctcctgtccagggaagctggccag 2416
          |||| | || ||| || ||| ||| |||||||||| ||       ||
Db  30509 AGCTACCGGCTGGTGCGAGCAGCCCCTCCTGACC---CACAGCGCAGACGCAGGACA 30565

Qy   2417 gggaagcctcagcagaaggcgggcccacacagagga-------gccacaggccaagaagg 2467
          || ||  || || ||||||| ||||| ||||||||        |||||||||||||||||
Db  30566 GAGAACCCCACACAGAGGAGGGGCCCCACACAGATGAACCACAGCCACAGAGCCAAGAAGG 30625

Qy   2468 tcagaggcccagtga 2481
          ||||| ||||||||||
Db  30626 TCAGAGCCCCAGTGA 30639
```

Right column

```
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="46F9-47A5"
                     /cell_line="cultured S2"
     gene            1..2690
                     /gene="JhI-1"
                     /note="28..2397"
     CDS             28..2397
                     /gene="JhI-1"
                     /note="JhI-1"
                     /codon_start=1
                     /product="juvenile hormone-inducible protein 1"
                     /protein_id="AAF99588.1"
                     /db_xref="GI:9802011"
                     /translation="MISANSPLFQFIRSPRLQTLTIRMVLVKSAGSPIYRTLRPLTTS
                     NLMAATIASAKDPLTGPRYEREPNVLRRKLASVVPGTVNLQVLCSGANGAPAAVVLFT
                     ADPILLAPROIKPWLEFVNRQIETVEDAYTLVCNGELLASPLSGEQVERLGITSIST
                     CLVRHCPNSFGISLTLAAKHNSEPVKIYLSGDDMPCQDLIDLGRDSTVLIHEATMEDD
                     LEEEARLKTHSTVSQAIQQGRNMANRHTILTHFSQRYAKCPRLFSDEDMQRVAIAFDN
                     MEVTIEDLQHYHKLYPALFAMYAEYTEELEQRAVKRELKQERKRLAET"

BASE COUNT     751 a    650 c    644 g    645 t
ORIGIN

Query Match     50.5%;  Score 133.4;  DB 5;  Length 2690;
Best Local Similarity   50.5%;  Pred. No. 5.4e-21;
Matches  475;  Conservative   0;  Mismatches  436;  Indels  30;  Gaps  5;

Qy   1438 taccagaaatcatcttccttggaacaggggtctgccatcccgatgaagattcgaaatgtc 1497
          |||| ||| | || || |||||| ||||| ||| | |  |||| ||| |||| ||| ||||
Db   1456 TACCCGAAGAATAATTTTTCTTGGGCACAGGGTCTCCTCCGCATTCCCAAATAAGACGGTG 1515

Qy   1498 agtgccacacttgtcaacataagcccgacacgtctctgctactggactgtggtggttgggc 1557
          || |||| | || |  ||    |  |  ||| |||| |||| || ||| | || ||| ||||
Db   1516 AGCTCCATTCTGATCAGGACTGCCAATTGATGCCTACGTGTTCGTTGGATTGTGGGAGAAGGA 1575

Qy   1558 acatttggcaagctgtgccgtcattacggaagaccag---gttggacaaggtcctgggcacc 1614
          || ||||| || ||| ||  || |||  |||||||| |     |||| | |||| ||| ||||
Db   1576 ACTTATGGTGCAAATTGTACGGACTATATGGACAGCGAGAAAGGGCAGTTGATTCTTCGGCCAG 1635

Qy   1615 ctggcgctggtcctgtccacacctgcaagdgtcacacacacgggcttgccaagtatc 1674
          || || | ||||| |||  | || | ||  |||| |||| |  ||||||||| |
Db   1636 CTACAGGCGATTTATGTATCCCATTGCATGCGACCACCACATTGGATTGATTGGTCTG 1695

Qy   1675 ttgctgcagagagaacgcgccttggcatctttgggaaagccgcttcacccctttgctggtg 1734
          || |||||||| || || ||| | |||  |  ||||  ||| |||  |||||||| ||| ||
Db   1696 CTCCGGGAAAG---------GAGGCAATTAAAACCTAGGCAGACCACTTATTCTG 1743

Qy   1735 gttgccccaaccagctcaaagcctggctccagaagtaccacacaacaagtgccaaggaggtc 1794
          || ||||||||||  || |||| |||||||||||||   || |||  || || ||||  |
Db   1744 CTGGCTCCTCCGCCAAAATTAAACCTTGGTT---GGAGTTCTTACAAATCGACAAAATAGAAACC 1800

Qy   1795 ctgcaccacatcagtatgattcctgccaaaatgcttcaggagggggctgagaatccagt 1854
          |||| |  ||  |  ||| ||| | |  ||| || || || |||| ||| ||| ||||
Db   1801 GTAGAGAGATCCGTATACTCTAGTGGGCAACGGTGAACTCCTTGCTAGTCCCCTGAGCGGT 1860

Qy   1855 cctgcagttggaaagatgatcagttcgctgttgcgaacatgtgattggaagagttcag 1914
          || || ||||| || ||| | ||   ||| || ||  |||| ||| | || |||| |||
Db   1861 GAACAAGTTGAACGTCTAGGAATTACGTCCATATCCACCTCCACTAGTAGGCACTCGCCC 1920

Qy   1915 acctgtctggtgcggcactgcaagacagtgttggctggctggtgcacacctggc 1974
          |||||| ||||  |||||||| || | ||| | |  |||| ||||| |||||||
Db   1921 AACTCTTTCGGAATAAGGCCTAACTCTGGCG------GCAAAGCACAATAGCCGAACCC 1971

Qy   1975 tggaaagtggtctattccggggacaccatgccttcggaggtctcgtcggatggggaaa 2034
          |||||||| |||  | |||||| ||  |||||   ||| |||  ||||| |||||  |||
Db   1975 ...
```

Bottom records (left column)

```
RESULT 11
LOCUS       AF215894    2690 bp    mRNA    INV    15-AUG-2000
DEFINITION  Drosophila melanogaster juvenile hormone-inducible protein 1
            (JhI-1) mRNA, complete cds.
ACCESSION   AF215894
VERSION     AF215894.1    GI:9802010
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2690)
  AUTHORS   Dubrovsky,E.B., Dubrovskaya,V.A., Bilderback,A.L. and Berger,E.M.
  TITLE     The isolation of two juvenile hormone-inducible genes in drosophila
            melanogaster
  JOURNAL   Dev. Biol. 224 (2), 486-495 (2000)
  MEDLINE   20387157
REFERENCE   2  (bases 1 to 2690)
  AUTHORS   Dubrovsky,E.B., Dubrovskaya,V.A. and Berger,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1999) Biology, Dartmouth College, Hanover, NH
            03755, USA
FEATURES             Location/Qualifiers
     source          1..2690
                     /organism="Drosophila melanogaster"
```

```
RESULT 12
LOCUS       AC017383     115873 bp      DNA         HTG        09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017383
VERSION     AC017383.1  GI:6553603
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 115873)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210513 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..115873
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    31582 a    25315 c    25855 g    33121 t
ORIGIN
```

```
Qy    1972 GTCAAGATCACTTACAGGGGCGACTATGCCATGCGACTGTCAGGATCGATCGATCTGGGCCGC 2031
           |||||||| |  ||| |||||
Db    2035 gatgccaccctcctgatacatgaaagcaccctggaagatggtttggaagaggaagcagtg 2094
Db    2032 GACTCCACCGTTCTTATTCACGAGGCGACAATCGAGGCGACCATTGGAGAGGAGGAGGCGCG 2091
Qy    2095 gaaaagacacacagcacacaacgtcccagcagccatcaagcgtggggatgaacgcggag 2154
           |||||| ||  ||||||||
Db    2092 TTAAAGACGACACAGCACCGTGTCGCAGGCTATTCAGTCAGGGGCGGGAATATGAACGCCCGC 2151
Qy    2155 ttcattatgctgaaccacttcagccagcgctatgccaagtgtccccctctt---cagccc 2211
           |||||| || | ||
Db    2152 CATACGAGTTCTCACCCACTTTTCGCAGGTTAGCGTTGCCCGCCGACTGAT 2211
Qy    2212 aacttcagcgagaaagtggggagttgccttttgacccacatgaaggtctgctttggagactt 2271
           |||| || | |||||| |
Db    2212 GAGGATAATGCAACGAGTTGCAAATAGCCTTTGATAATATGGAGGATAAGCCATTGAGGATCTG 2271
Qy    2272 ccaacaatgccaaagctgattccccactgaaagcccctgtttgctggcgaacatcgaggag 2331
           |||| | ||| | ||||||
Db    2272 CAGCACTACCACAAGGTCTACCCCGCCCCTTTTCGGATGTACGCCCGAGTACACGGAAGAA 2331
Qy    2332 atggaggagcgcaggggagaagcgggagctgcgcaggtgcg 2372
           |||||||||
Db    2332 CTGGAGCAGGCAGCGGCGGCTGTCAAGCGCGGGGAGCTGAAACAGGAGCG 2372
```

```
RESULT 13
LOCUS       AC007417     175682 bp      DNA         INV        02-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 47A-47B, BAC clone
            BACR48F07, complete sequence.
ACCESSION   AC007417
VERSION     AC007417.4  GI:13184056
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 175682)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Ananaides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Gocayne,J.D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
```

```
Db   63779 ACTTATGGTCAAATTGTACGACTATATGGACGACGAGAAAGGGCAGTTGATTCTTGCCAG 63838
Qy    1615 ctggctgctgtgtttttgtgtcccaccctgcacgcagatcaccacgggcttgccaagtatc 1674
           ||  | |   ||||
Db   63839 CTACAGGCGATTTATGTATCCCATTTGCATCCGACCACCACATTGGATTGATTGGTCTG 63898
Qy    1675 ttgctgcagagagaacgcgcctttgcaatctttttggaaatggacgcttcaacccttgctggtg 1734
           || |  |||| ||||
Db   63899 CTCCGGGAAAG------GAGGCAATTAAAACCTAGGCGACACCCACTTATTCTG 63946
Qy    1735 gttgccccaaccagctcaaagcctggctccagcagtgtaccacaccaagtgccaggagtc 1794
           |||| |||| ||
Db   63947 CTGGGCTCCTGCCCAAATTGAACCTTGGTT---GGAGTTCTACAAATCGACAAAATAGAAACC 64003
Qy    1795 ctgcaccacatcagtatgatccctgccaaatgccttcaggaagggggctgagatctccagt 1854
           || ||||  ||||||||
Db   64004 GTAGAAGATGCGTATACTCTAGTGGGCAACGGTGAAGCTGAACTCCTTAGCTAGTCCCTGACGGT 64063
Qy    1855 cctgcagtggaagaagattgatcaagttcgctctgttgccaacatgtgattgggaagtttcag 1914
           |||||| |||||||
Db   64064 GAACAAGTTGAACGTCTAGGAATTACGTCCAATATCCACCTCGCCTCCTAGCACTGCCCCC 64123
Qy    1915 acctgtctgtgtgcgggcactgcaagcaatgcgtttggctgtcgcgtcgtgtgcacacctctggc 1974
           |||| ||||
Db   64124 AACTCTTTCGGAATAAGCCTAAACTCTGGCC------GCAAACACAATAGCCAAACC 64174
Qy    1975 tggaaagtggtgctctattccgggacacacatgccctgcgcgaggtctcgtgttccggatggggaaa 2034
           |||  ||||||
Db   64175 GTCAAGATCACTTACAGCGCGACCACTATGCCATGTCAGGATCGATCGATCTGGGCCGC 64234
Qy    2035 gatgccaccctcctgatacatgaaagcaccctggaagatggtttggaagaggaagcagtg 2094
           |  | |  ||||
Db   64235 GACTCCACCGTTCTTATTCACGAGGCGACAATCGAGGCGACCATTGGAGAGGAGGAGGCGCG 64294
Qy    2095 gaaaagacacacagcacacaacgtcccagcagccatcaagcgtggggatgaacgcggag 2154
           |||||| ||  ||||||||
Db   64295 TTAAAGACGACACAGCACCGTGTCGCAGGCTATTCAGTCAGGGGCGGGAATATGAACGCCCGC 64354
Qy    2155 ttcattatgctgaaccacttcagccagcgctatgccaagtgtccccctctt---cagccc 2211
           |||||| || | ||
Db   64355 CATACGAGTTCTCACCCACTTTTCGCAGGTTAGCGTTGCCCGCCGACTGAT 64414
Qy    2212 aacttcagcgagaaagtggggagttgccttttgacccacatgaaggtctgctttggagactt 2271
           |||| || | |||||| |
Db   64415 GAGGATAATGCAACGAGTTGCAAATAGCCTTTGATAATATGGAGGATAAGCCATTGAGGATCTG 64474
Qy    2272 ccaacaatgccaaagctgattccccactgaaagcccctgtttgctggcgaacatcgaggag 2331
           |||| | ||| | ||||||
Db   64475 CAGCACTACCACAAGGTCTACCCCGCCCCTTTTCGGATGTACGCCCGAGTACACGGAAGAA 64534
Qy    2332 atggaggagcgcaggggagaagcgggagctgcgcaggtgcg 2372
           |||||||||
Db   64535 CTGGAGCAGGCAGCGGCGGCTGTCAAGCGCGGGGAGCTGAAACAGGAGCG 64575
```

```
Query Match   5.4%;    Score 133.4;   DB 64;   Length 115873;
Best Local Similarity   50.5%;   Pred. No. 3.8e-21;
Matches   475;   Conservative   0;   Mismatches 436;   Indels 30;   Gaps 5;

Qy    1438 tacccagaaatcatcttccttggaacaaggggtctgccatccccgatgaaagattcgaaatgtc 1497
           ||||| | ||||  |  ||| ||  |||||  ||| |    | ||||||||
Db   63659 TACCCGAAGATAATTTTCTTGGGCACAGGCCTCCTGCATTCCCAATAAGACGGAACACGTG 63718
Qy    1498 agtgccacacttgtcacatagaccccacagctctctgcactggactgtggtgaaggc 1557
           |||  |||  |||| ||
Db   63719 AGCTCCATTCTGATCAGACTGCAATTGATGCCTACGTGTTGTTGGATTGTGTGGAGAAGGA 63778
Qy    1558 acaatttgggcagctgtgccgtcattacggagaccag---gtggacaaggtcctgggcacc 1614
           |||  ||  | |  |   ||       ||   ||    |     | |||  ||   ||
```

```
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,D., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Sviskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 47A-47B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175682)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,J.,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Sviskas,R.R., Wan,K.H., Weiburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 2, 2001 this sequence version replaced gi:5670592.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp&fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..175682
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="47A-47B"
                     /clone="BACR48F07 (D625)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    49836 a   38215 c   38034 g   49597 t
ORIGIN

Query Match     5.4%;  Score 133.4;  DB 4;  Length 175682;
Best Local Similarity   50.5%;  Pred. No. 3.6e-21;
Matches  475;  Conservative    0;  Mismatches  436;  Indels   30;  Gaps    5;

Qy   1438 taccagaaatcatcttccttggaacaaggttctgccatccgatgaagaatgtc 1497
          |||  | |||  ||||||||| || ||||| ||||| ||||| || || |||||
Db 108897 TACCCAAAGATAAATTTTCTTGGGCACAGGGCTCCTGCAATTGAACATCGAAACGTG 108838

Qy   1498 agtgccacacttgtcaacataagcccgaacag---gtggacaaggtcctgggccacc 1557
          || |||| || ||| ||| || | |||| | ||   |||||| ||||| | |||||
Db 108837 AGCTCCATTCTGATCAGGACTGCAATTGAACGCCTACGTGTTGTTGTGGAGAAGGA 108778

Qy   1558 acattgggtcaagctgccgtcaatacgggaccag---gtggacaaggtcctgggccacc 1614
          || | ||||| | | |||  ||||| | ||| ||   ||||||||||| | |||| ||||
Db 108777 ACTTAYGGTCAAATTGTACGACTATATGGCACACGAAAAGGCAGTTGATTGCTTCGCCAG 108718

Qy   1615 ctggctgctgttttgttgtgtccccacctgcacgcacggcaagtatc 1674
          ||  | || |||| ||| | ||||||||  || ||||  || |||
Db 108717 CTACAGCGGCGATTTATGTATCCCATTNCATCCGGACCACATTGGATTGGTCTG 108658

Qy   1675 ttgctgcagagaagaacgcgcttgggccatcttccacccttttgctggtg 1734
          ||||| ||||||||||| || ||||||||||| || |||||||||||| |
Db 108657 CTCCGGGAAAG------------GAGGCAATTAAAACCTAGGCAGGACCCACTTATTCTG 108610

Qy   1735 gttgccccaaccagctcaaagcctggctggctccagagtaccacacaaccagagtcgaggtc 1794
          |||||||| || ||| |||||| ||| | |||| | || ||| |  |   ||| |||
Db 108609 CTGGCTCTCGCCAAATTGAACCTTGGTT---GGAGTTCTACAATCGACAAAATAGAAACC 108553

Qy   1795 ctgcaccacatcagtatgattcctgccaaatgccttcaggaagggggctgagactccagt 1854
          ||||  ||| || |||| || | ||| ||||| ||||| ||||||| || || |||||
Db 108552 GTAGAAGATGCGTATACTCTAGTGGGCCAACGGGTGAACTCCTAGCTAGTCCCCTGAGCGGT 108493

Qy   1855 cctgcagttggaagattgatcagttcgctgttggcgaacatgtgatttggaaagttcag 1914
          || ||||| ||||||| ||||||| || |||| ||  ||| || || ||| |||| |||
Db 108492 GAACAAGTTGAACCTTCTAGGAATTACGTTCCATTANCCACCTGCCTAGTTAGGCACTGCCC 108433

Qy   1915 acctgtctggttgcgggcactgcaagcatgcgtttggctgtgttgcacacctctggc 1974
          ||||| || | ||| | ||||| ||| | |||||||| | |  |||||| | |||
Db 108432 AACTCTTTCGGAATAAGCCTAACTCTGGCG------GCAAAGCACAATAGCGACACCC 108382

Qy   1975 tggaaagtggtcattccgggacaccatgccctgcgaggtctggttccggatgggggaaa 2034
          ||||||||| || | ||||||| |||| ||||| ||||||| | |||||| ||| ||
Db 108381 GTCAAGAATCACTTACAGGCGGGCCGACACTATCCAGTGCCTATGGATCTGATCGANTCGGGCC 108322

Qy   2035 gatgcaccctctgatacatgaagcaccccctggaagaatggtttggaagaagacagtg 2094
          ||||| | ||||||| | || |||| | ||||||| |||||||||| |||| ||| ||
Db 108321 GACTCCACCGTTCTTATTCACGAGGCGACAAATGGACAGCTTGGAGCGAGGAGGAGGCGCG 108262

Qy   2095 gaaaagacacacgcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggag 2154
          |||||||||| | | || ||||||||| | | ||| ||||| ||| | ||| |||||
Db 108261 TTAAAGAACGGCACAGCACCACCGTTGTCGCAGGCTATTCAGCNCGGGGCCGGGAATATGAACGCCCGC 108202

Qy   2155 ttcattatgctgaaccacttcagccagcgctatcgcaggtaccccctctt---cagcccc 2211
          || | ||| || |||||| ||| |||| ||| | |  || |||||| |   ||| ||
Db 108201 CATACGAATTCTCCCACTTTTCGGCAGGTTACGTTACGTCGCCCGGATTACCGAGTGAT 108142

Qy   2212 aacttcagtcgagaaaagtgggagttgccttttgaccacatgtctgctttgggagactt 2271
          || ||||| |||| ||| ||||||| |||| | ||| |||| |||| || | |||||
Db 108141 GAGGATATGCAACGAGTGCAAGTTGCCATTGAGAATGGAAGTGACAGCGAAGGTCTTGTTGAGGAGTCG 108082

Qy   2272 ccaacaatgcccaagctgattccccactgaaagccctgtttgctggcgacacatggaggag 2331
          || || | || |||| | || |||| | ||||| ||||| ||||| |||| ||| ||||
Db 108081 CAGCACTACCACAAAGCTCTACCCCCGTTTCGCGATGGTACGCCCGAGTACACAGGCGGAGAA 108022

Qy   2332 atggaggaggcgcaggggaggagcggggagctgcggcgagggtgcg 2372
          ||||| ||||| ||| |||||||| || |||||| | || |||
Db 108021 CTGGAGCAGGCGGGCGCTGCTCAAGCGGGAGCTGAAACAGGACG 107981

RESULT 14
AC007352
AC007352

LOCUS       AC007352     190574 bp    DNA     HTG       24-FEB-2000
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPCI-98
            19.J.16 map 47A-47B strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 36 unordered pieces.
ACCESSION   AC007352
VERSION     AC007352.8  GI:7025687
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 190574)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,V., Hoskins,R.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,J.,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Sviskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190574)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
```

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hunmasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,M., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.

TITLE   Direct Submission
JOURNAL   Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley
          Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT   On Feb 24, 2000 this sequence version replaced gi:7018750.
          For further information about this sequence, including its location
          and relationship to other sequences, please visit our sequence
          archive Web site (http://www.fruitfly.org/sequence/) or send email
          to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
          the following cutoffs: length >= 200 bases.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 36 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *   1   678: contig of 678 bp in length
          *   679   758: gap of unknown length
          *   759   1358: contig of 600 bp in length
          *   1359   1438: gap of unknown length
          *   1439   2373: contig of 935 bp in length
          *   2374   2453: gap of unknown length
          *   2454   2962: contig of 509 bp in length
          *   2963   3042: gap of unknown length
          *   3043   3889: contig of 847 bp in length
          *   3890   3969: gap of unknown length
          *   3970   5437: contig of 1468 bp in length
          *   5438   5517: gap of unknown length
          *   5518   6572: contig of 1055 bp in length
          *   6573   6652: gap of unknown length
          *   6653   8387: contig of 1735 bp in length
          *   8388   8467: gap of unknown length
          *   8468   9060: contig of 593 bp in length
          *   9061   9140: gap of unknown length
          *   9141   11167: contig of 2027 bp in length
          *   11168   11247: gap of unknown length
          *   11248   14518: contig of 3271 bp in length
          *   14519   14598: gap of unknown length
          *   14599   18837: contig of 4239 bp in length
          *   18838   18917: gap of unknown length
          *   18918   24324: contig of 5407 bp in length
          *   24325   24404: gap of unknown length
          *   24405   29984: contig of 5580 bp in length
          *   29985   30064: gap of unknown length
          *   30065   35794: contig of 5730 bp in length
          *   35795   35874: gap of unknown length
          *   35875   44808: contig of 8934 bp in length
          *   44809   44888: gap of unknown length
          *   44889   58222: contig of 13334 bp in length
          *   58223   58302: gap of unknown length
          *   58303   78322: contig of 20020 bp in length
          *   78323   78402: gap of unknown length
          *   78403   122249: contig of 43847 bp in length
          *   122250   122329: gap of unknown length
          *   122330   179102: contig of 56773 bp in length
          *   179103   179182: gap of unknown length
          *   179183   179407: contig of 225 bp in length
          *   179408   179487: gap of unknown length
          *   179488   180592: contig of 1105 bp in length
          *   180593   180672: gap of unknown length
          *   180673   181383: contig of 711 bp in length
          *   181384   181463: gap of unknown length
          *   181464   182092: contig of 629 bp in length
          *   182093   182172: gap of unknown length
          *   182173   183480: contig of 1308 bp in length
          *   183481   183560: gap of unknown length
          *   183561   183771: contig of 211 bp in length
          *   183772   183851: gap of unknown length
          *   183852   184058: contig of 207 bp in length
          *   184059   184138: gap of unknown length
          *   184139   185035: contig of 897 bp in length
          *   185036   185115: gap of unknown length
          *   185116   185846: contig of 731 bp in length
          *   185847   185926: gap of unknown length
          *   185927   186560: contig of 634 bp in length
          *   186561   186640: gap of unknown length
          *   186641   187253: contig of 613 bp in length
          *   187254   187333: gap of unknown length
          *   187334   187611: contig of 278 bp in length
          *   187612   187691: gap of unknown length
          *   187692   188409: contig of 718 bp in length
          *   188410   188489: gap of unknown length
          *   188490   189196: contig of 707 bp in length
          *   189197   189276: gap of unknown length
          *   189277   189969: contig of 693 bp in length
          *   189970   190049: gap of unknown length
          *   190050   190574: contig of 525 bp in length.
FEATURES   Location/Qualifiers
     source   1..190574
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="47A-47B"
                     /clone="BACR19J16 (D626) RPCI-98 19.J.16"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"
BASE COUNT   53370 a   39726 c   40358 g   54299 t   2821 others
ORIGIN

Query Match   5.4%; Score 133.4; DB 60; Length 190574;
Best Local Similarity 50.5%; Pred. No. 3.6e-21;
Matches   475; Conservative   0; Mismatches   436; Indels   30; Gaps   5;

```
Qy     1438 taccagagaaatcatcttccttggaacaggtgtctgccatcccgatgaagatttcgaaatgtc 1497
            ||||  ||| ||   ||   ||||| |||| | | | | ||| || |   ||| ||||  | |
Db   132812 TACCCGAAGATAAATTAATTTTCTTGGGCACAGGCCTCCTGCCATTCCCAATAAGACGCGGAAACGTG 132871

Qy     1498 agtgccacacttgtcaacataagcccgacaacgtctctgctactggactgtgttgagggc 1557
            ||| ||||   ||   ||   ||||||   || |  ||  ||| |  |||| | | |
Db   132872 AGCTCCATTCTGATCAGGACTGCAATTGATGCCTACGTGTTGTTGGATTGTGGAGAAGGA 132931

Qy     1558 acatttgggcagctgtgcgcgtcattacggagaccag---gtggacaggtgtcctgggcacc 1614
            || |||||  || ||||   | |  |  | || || |    || ||| |||  |||| |
Db   132932 ACTTATGGTCAAATTGTCAACTATATGGACACGAGAAAGGCAGTTGATTCTTCCCCAG 132991

Qy     1615 ctggctgctgtgtttggtgtgtccacctgcacgcagatcaacacacgggcttgccaagtatc 1674
            ||| || | | | ||| | | || |||  |  |  ||   || ||| || || ||||  | |
Db   132992 CTACAGCCGGATTTATGTATGTCCCCATTGCATGCCGACCACCACAGTTGGATTGCATTGGTCTCG 133051

Qy     1675 ttgctgcagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtg 1734
            || ||| | | |||  || ||  ||  ||| ||||||| ||||   |  ||||| ||| |
Db   133052 CTCCGGGAAAG------------GAGGCAATTAAAACCTAGGGCAGCAGACCCACTTATTCTG 133099

Qy     1735 gttgccccccaaccagtcaaagcctggcctccagcagtaccacacaccagccaggagggtc 1794
            |||||| || ||  ||| ||||||||  | ||     ||| ||||||  ||| |    | |
Db   133100 CTGGCTCCTCGCCAAATTGAACCTTGGTT---GGAGTTCTACAANCGACAANATAGAAAC 133156

Qy     1795 ctgcaccaatcagtgatgattcctgccaaatgtcttcaggaagggggctgagatctccagt 1854
            |||| || | ||| |  | |||   || |||  |  | || ||||||| || |||  |||
Db   133157 GTAGAAGAATCCGTATACTCTAGTGGGGCAACGGTGAACTCCTAGCTCCGTGAGCGGT 133216

Qy     1855 cctgcagtcgaaaagattgatcagttcgctgttgcgaacatgtgattggaagaagtttcag 1914
            || ||| || | |||||| ||||| ||| |||| ||||||| || ||| |||||| |||
Db   133217 GAACAAGTTGAAACGTCTAGGAATTACGTCCATATATCCACCTGCCTAGTTAGGCACTGCCCC 133276
```

```
Qy    1915 acctgtctggtcggcactgcaagcatgcgtttggctggtcggctggtgtcacacctctggc 1974
           ||||||| |||||||| |||||  |||||||||||||||| |||      .
Db  133277 AACTCTTTCGGAATAAGCCTAACTCTGGCG---------GCAAAGACAATAGCGAACC 133327

Qy    1975 tggaaagtgtctattccgggggaaccaatgcctgccgaggtctctgccctgatggggaaa 2034
           || ||| ||||||||  |||||| |||||||||||||||||||||||||||| ||||
Db  133328 GTCAAGAATCACTTACACGGCGGCCGACACTATGCCATGTCAGGAATCTGATCGAATCTGGGCGC 133387

Qy    2035 gatgccaccctcctgatacatgaagccaccctggaagatgggtgttggaagaaggaagtg 2094
           || |||||| |||| |||||||||||| ||||||||||||| |||||  .
Db  133388 GACTCCACCGTTCTTATTCACGGAGCCACAATGGACGGAGGATGACTTGGAGGAGGGCCGG 133447

Qy    2095 gaaaggacacacgcacaacgtcccaacgcaatggtcccaacgcaacggggag 2154
           ||||||||| |  |||||||||||  ||||||||||   |||||||||||
Db  133448 TTAAAGACGCAACAGCACCACCGGTGCGCACGGCTAATCACGACAACGACCCCGC 133507

Qy    2155 ttcattatgctgaaccacttcagccagcgcgtatgccaaggtcccctctt---cagccc 2211
           ||||  ||||||||||||||||||||  ||||||||||||||||||  ||
Db  133508 CATACGATTCTCTCCCACTTTCGCACCGTTACGCTAAGTGCCCGGCAATTACCGAGTGAT 133567

Qy    2212 aacttcagcgagaaagtgggaagttgcctttgaccacaatgaaggtctgcgtttggagactt 2271
           ||||||||||||| |||| |  ||||||||| |||||||  |||||||||||||| ||
Db  133568 GAGGAATATGCAACGAGTTGCAATAGCCTTTGATTAATATGGAAGTGACCGTTGAGGAT CTG 133627

Qy    2272 ccaacaatgccaagctgattccccactcagccagccccctcctt---cagccc 2331
           ||| ||||||||||||||||  ||||||| |||||||||||| ||
Db  133628 CAGCACTACCACAAGCTCTACCCCGCCCTTTTCGGATTGTACGCCGAGTACACGGAAGAA 133687

Qy    2332 atggagagcgaggaagcggggagaaagctgcggcaggtgcg 2372
           ||||||||||||||||||||||||||||||||||||||| |||
Db  133688 CTGGACGAGCGGCGCGGCTGTCAAGCGGGGACTGAAACAGGAGCG 133728
```

RESULT 15

FEATURES   Location/Qualifiers
     source   1..261846
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA   join(960..1069,1135..1237,1290..1357,1416..1731,
                     2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,
                     6499..6941,6999..7841,7921..8315)
                     /gene="pfk"
                     /product="CT13302"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     gene   <960..>8315
                     /gene="pfk"
                     /note="pfk gene product [alt 1]"
                     /codon_start=1
                     /map="4E3-4E6"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     CDS   join(1026..1069,1135..1237,1290..1357,1416..1731,
                     2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,
                     6499..6941,6999..7841,7921..8059)
                     /gene="pfk"
                     /product="CT13302"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
                     /protein_id="AAF58840.1"
                     /db_xref="GI:7303792"
                     /translation="MHSIKFRVSFTKLKPFLEINGRIPICRHHGPTTFRLEISNKTP
                     PIRQKLTFPNIGLTQCTRSHHLCCPRDISGNTLLSVKFNCKRHCTKLRSDSGDQKNDSP
                     GERNIQKDKSAQRCGKPINLNHGFLNAVNYSRKNAVRKVKKKSAPKRCGKSVDELRKC
                     RVPLMECVERTQAVAKAMAEKRWADAVKLRGRSFERNLETYKMLTRLKPPKENFDADG
                     KGIEGYRLAVMHIGAPACGMNAAVRSFVRNAIYRGDVVYGINDGVEGLLIAGNVRELGW
                     SDVSGWVQGGAVLGTKRTLPEGKFKEIAARLKEFKIQGLLLIGGFESYHAACQIADQ
                     RDNYPQFCIPIVVIPSTISNNVPGTEFSLGCDYGLNEITEICDRIRQSAQGTRRVFV
                     IETMGGYCGYLATIAGLAGGADAAYIYEEKFSIKDLQQDVYHMASKMAEGVSRGLIIR
                     NEKASENYSTDFIYRLYSEEGKGLFTCNRLILHMQQGGSPTPFDRNMGTKMAAKCVD
                     WLAAQIKANIDANGVNVNCKSPDFATLLGIVSROYRFSPLVDLIAETNFDQRIPKKQWW
                     LRLPLLRILAKHDSAYEEGMITTVEEECDTDAVA"
     mRNA   join(2152..2759,3801..4142,4200..4354,4727..4835,
                     6101..6289,6499..6941,6999..7841,7921..8315)
                     /gene="pfk"
                     /product="CT13275"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"

```
     CDS          join(2613..2759,3801..4142,4200..4354,4727..4835,
                  6101..6289,6499..6941,6999..7841,7921..8059)
                  /gene="Pfk"
                  /note="Pfk gene product [alt 2]"
                  /codon_start=1
                  /db_xref="FLYBASE:FBgn0004001"
                  /gene="CG11867"
                  /db_xref="FLYBASE:FBgn0003071"
                  /protein_id="AAF58841.1"
                  /db_xref="GI:7303793"
                  /translation="MNEINQRFLARGSQKDKGLAVFTSGGDSQGMNAAVRACVRMAI
                  YLGCKVFPIREGYQGMVDGGDCIQEAWNASVSSIIHRGGTIIGSARCQDFRERQGRLK
                  AANNLIQRGITNLVVIGGDGSLTGANLFRQEWSSLLDELVKNKTITVEQEKRNVLHI
                  LVGGLACEADFIFIPEMPPKVDMPDRLCSQLAQERSAGQRLNIVIVAEGAMDREGHPI
                  TAEDVKVIDERLKHDARITVLGHVQRGGNPSAFDRILACRMGAEATIALMEATKDSV
                  PVVISLDNQAVRVPLMECVERTQAVAKAMAEKRWADAVKLRGRSFERNLETYKLMLTR
                  LKPPKENFDADCKGIEGYRLAVMHIGAPACGMNAAVRSFVRNAIYRGDVVYGINDGVE
                  GLIAGNVRELGWSDVSGWGWQGGQGAYLGTKRTLPEGKKFKEIAARLKEFKIQGLLIIGGF
                  ESYHAAGQIADQRDNYFQFCIPIVVIPSTISNNVPGTEFSLGCDTGLNEITTEICDRIR
                  QSAQGTKRRVFVIETMGGYCGYLATLAGLAGGADAAYIYEEKFSIKDLQQDVVHMASK
                  MAEGVSRGLLRNEKASENYSTDFIYRLYSEEGKGLFTCRMNILGHMGQGGSPTPFDR
                  NMGTKMAAKCVDWLAAQIKANIDANGVVNCKPGDTAILLGIVSRQYRFSPLVDLIAET
                  NFDQRIEPKKQWWLRLPLLRILAKHDSAYEEGMYIIVEEECDTDAVA"

     mRNA         join(8915..9052,9216..9726,9803..
                  11235..11542,11596..>12210)
                  /gene="CG11866"
                  /product="CT37014"
                  /db_xref="FLYBASE:FBan0011866"
                  /gene="CG11866"
                  /db_xref="FLYBASE:FBgn0033486"

     gene         <8915..>12210
                  /gene="CG11866"
                  /db_xref="FLYBASE:FBan0011866"
                  /gene="CG11866"
                  /db_xref="FLYBASE:FBgn0033486"

     CDS          join(9286..9726,10663..10911,10976..11174,11235..11542,
                  11596..12210)
                  /gene="CG11866"
                  /note="CG11866 gene product"
                  /codon_start=1
                  /db_xref="FLYBASE:FBan0011866"
                  /map="46E4-46E6"
                  /db_xref="FLYBASE:FBgn0033486"
                  /protein_id="AAF58839.1"
                  /db_xref="GI:7303791"
                  /translation="MVSRQMCSGGGTSASEGYSPEGAGSGSASGTEMASGSAVSSG
                  SGTGVQEPPESTANRSSVLRRTSYQGNQAHSSYRVVSAGSLAESASAVTRLHSYTMQ
                  MQRQPFATTTTTTTSAPSLIYLDLPNELIEKILSYDVKVSNLRLVSHRMNDICMAML
                  NTAFTKQIKTTLSRFQATKASMPRESRRNHPLACEDIIETCYRMLSLLQMSMGKH
                  IERGHCCFPGAILDEVQAILNYISITPRLQRPYRVTDELFDLSTMAMEYFKDRIEAT
                  LPGLAYFNKDFYTLPTTTKRPFLAISDEDSASNSPPQNHWLVRLGIRKIKQGMKMY
                  NNQLSVLRTELRRTCKKRKAEQSKQLAEQSNLLSEQQKQTLEVANRLDENDKRNEEMSR
                  KFSTLLQEELNCKTELQQFWRSKSPAIPAVCSSCNQKVAPVVPPEDFQOALVNQGVDPEH
                  FIIINEDVDAESDVSGELKEFASPDSTTAKILLAVSTAAKNLRKKMPSESQSNAIAS
                  TSKAAEVAQSOSOPMATGNAGAANKAAAGSYSPKLFYGNHQRSGVIVSPVSQNLGVAATT
                  DNVAPGSSEALEEPEEAKKARRVQKANRYVNTPGKRSK"

     mRNA         complement(join(<12115..>.12937,12999..13642,13694..13708))
                  /gene="CG17753"
                  /product="CT11457"
                  /db_xref="FLYBASE:FBan0017753"
                  /db_xref="FLYBASE:FBgn0033487"
                  complement(<12115..>13708)

     gene         complement(<12115..>13708)
                  /gene="CG17753"
                  /db_xref="FLYBASE:FBan0017753"
                  /map="46E4-46E6"
                  /db_xref="FLYBASE:FBgn0033487"

     CDS          complement(join(12802..12999,12937,12999..13642,13694..13708))
                  /gene="CG17753"
                  /note="CG17753 gene product"
                  /codon_start=1
                  /db_xref="FLYBASE:FBan0017753"
                  /db_xref="FLYBASE:FBgn0033487"
                  /protein_id="AAF58838.1"
                  /db_xref="GI:7303790"
                  /translation="MSSIKIEFAVQMRRGDESYAGALRSALDGVGVQVEIDIQEGRVII
```

```
                  QTQRPWSEIQDKIEATGVRAVLSGFGGQSAVALINTTGSVVDKTPIQGVVRFTTITAD
                  KKPGVVVDGVVDGLSPGLHGLHIHESGDYSAGCSSVGEHYNPRQSPHGSPAAGAEERH
                  AGDLGNIRADENGRATFREVPDVLEVWDIIGRAVVLTANADDLGRGGNDQSLIDGNSG
                  ERIACGIIARSAGILENFKRICACDGVTLWDERNKPLAGKERSQKL"

     mRNA         join(<18636..18761,19765..19876,19968..20213,22806..22878,
                  22958..23344,23489..23633,23934..24407,25026..>25202)
                  /gene="CG11867"
                  /product="CT37016"
                  /db_xref="FLYBASE:FBan0011867"
                  /gene="CG11867"
                  /db_xref="FLYBASE:FBgn0033488"
                  <18636..>25202
                  /gene="CG11867"

     gene         <18636..>25202
                  /gene="CG11867"
                  /map="46E4-46E8"
                  /db_xref="FLYBASE:FBan0011867"
                  /db_xref="FLYBASE:FBgn0033488"

     CDS          join(18636..18761,19765..19876,19968..20213,22806..22878,
                  22958..23344,23489..23633,23934..24407,25026..25202)

Query Match     5.4%;  Score 133.4;  DB 5;  Length 261846;
Best Local Similarity 50.5%;  Pred. No. 3.5e-21;
Matches  475;  Conservative  0;  Mismatches  436;  Indels  30;  Gaps  5;


Qy    1438 tacccagagaaatcatcttcctggaacacagggtctgccatcccgaatgaaagattcgaaatgtc 1497
           |||||   ||| | ||||| ||||| |||||| |||| ||||| ||| ||||  |  ||  | ||
Db  199291 TACCCGAAGATAAATTTTCTTGGGCACAGGTGTCCTCCATTCCCAATAAGACGGCGAAACGTG 199232

Qy    1498 agtgccacacttgtcaacataagcccgacaacgtctctgctactggactgtgtggaggc 1557
           ||| ||| | || || |  ||| || ||||| ||| |||| ||||  | | ||| |
Db  199231 AGCTCCATTCTGATCAGGACGTCAAATTGATGCCTCACGTTGTTGTTGGATTGTTTG 199172

Qy    1558 acattgggtcagctgtgccgtcattacgagagaccag---gtggacaggtcctgggcacc 1614
           ||  |||||||  ||||  | |  ||| |  ||| |||   || ||| ||| ||| ||
Db  199171 ACTTATGGTCAAATTGTACGACTATATGGACACGAGAAAGGGCAGTTGATTCTTCGCCAG 199112

Qy    1615 ctggctgtcgtgtttgtgtccccacctgcacgcagagatcaccacacacgggcttgccaagtatc 1674
           || | ||  | | |||| | ||||  ||||| ||||  ||| || ||||||  | |||| | |
Db  199111 CTACAGGCGGATTTATGTATCCCATTTGCATGCCGACCACACCATTGGATTGATTGGTCTG 199052

Qy    1675 ttgctgcagagagaacgcgccttggaatctttgggaaagccgcttcacccctttgctgttg 1734
           || ||  || |   || | || |||   | |||||||| |||  | ||||| |||||
Db  199051 CTCCGGGAAAG-------GAGGCAATTAAAACCTAGGGCAGACCCCACTTFATTCTG 199004

Qy    1735 gttgccccaaccagctcaaaagcctggctccagcagtacacacaacagtgccagaggtc 1794
           | |||  | |||   |  |||  ||||  | ||||| |||| ||| | || |||| | |
Db  199003 CTGGCTCCTCGCCCCAAAATTGGAACCTTGGTT---GGAGTTCTACAATCGACAAATAGAAAACC 198947

Qy    1795 ctgcaccacatcagtatgatcctgccaaatgcctcaggaaggggctgagactccagt 1854
           || ||  |  || |  |||| | ||||||||| | |||| ||||  || |||| ||
Db  198946 GTAGAAGATCCGTATACTCTTAGTGGGCACGGTGAACTCCTTAGCGTAGTTCCCCTGACCGGT 198887

Qy    1855 cctgcagtggaaagattgatcagttcgctgttgcggaacatgtgattgggaagtttcag 1914
           ||| || ||||   ||||| ||| || || || ||||||||| ||| ||||| ||| |||
Db  198886 GAACAAGTTGAACGTCTAGGAATTACGTCCATAATCCACCTGCCTAGTTAGGCACTCCCC 198827

Qy    1915 acctgtctgttgcggcactgcaagcagtcgtttgctgtgcgtggtgcaacctctggc 1974
           | ||| ||  ||  | ||| |||| | || ||| |||| ||||||||  |||| |||
Db  198826 AACTCTTTCGGAATAAGCCTAAACTCTGGCG------GCAAAGCACAATAGGGAACCC 198776

Qy    1975 tggaaagtggtctattccgggacacacatccccgcgaggcttctggtccggatggggaaa 2034
           || |||||| ||| ||| || ||||  ||| || |  ||  ||| |||| |||| |||
Db  198775 GTCAAGATCACTTACAGCGCGACACTATCCCATTCAGDATCTGATCGATCTGGGCCGCC 198716

Qy    2035 gatgccaccctcctgatacatgaagcgacacctggaagatggttggaagaagaagtg 2094
           ||||| ||||| || | ||| ||||| | ||| ||||| ||||| |||| || || ||
Db  198715 GACTCCAACCCTTCTTATTCACGAGAGCCAGCGGAGAATGAGACTTGGAGGAGAGCCGG 198656

Qy    2095 gaaaagagacaacacaagcacacaagtgccaacatcacgcgtggggagtgacgggatgaaacggag 2154
           | |||| ||| || || || || ||| |||| || ||| | ||||| |||| || |||| |||||
Db  198655 TTAAAGACGGCACAGCACCGGTGTCGACCAGGAAGGGCCGGGAATATGAAACGCCGC 198596

Qy    2155 ttcattatgctgaaccacttcagccagccagcgctatgcccaaggtcccccctctt---cagccc 2211
           || ||||||| ||||| || |||||||| || ||| | || |||| | ||| |||   |||||
Db  198595 CATACGATTCTCTCCCCACTTTTCCGCCAGCCGTTACCGTAGAGTGAT 198536
```

```
Qy   2212 aacttcagcgcgagaaagtgggagttgcctttgaccacatgaagtctgctttggagactt 2271
          |  |||||||  ||| |||||||  ||||||||  ||| |||| || ||  || |  |
Db 198535 GAGGATATGCAACGAGTTGCAATAGCCTTTGATAATATGGAAGTGACCGTTGAGGATCTG 198476

Qy   2272 ccaacaatgcccaagctgattcccccactgaaagccctgtttgctggccgacatcgaggag 2331
          |  ||||||  ||| |||  ||| ||| ||  || ||  || ||| |  ||  || ||
Db 198475 CAGCACTACCACAAGCTCTACCCCGCCCTTTTCGCGATGTACGCCCGAGTACACGGAAGAA 198416

Qy   2332 atggaggagcgcaggggagaaagcgggggagctggcggcaggtgcg 2372
          |||||| || ||||||||  |||||| ||| |||||| |||
Db 198415 CTGGAGGCAGCGCGGGCTGTCAAGCGGGAGCTGAAACAGGGAGCG 198375
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: October 28, 2001, 18:04:11 ; Search time 163.2 Seconds
                                            (without alignments)
                                            9545.490 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgcttgctccgct...........agaaggtcaagagccagtga 2481

Searched:
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Post-processing: Minimum Match 0%
Minimum DB seq length: 0
Maximum DB seq length: 200000000
               Maximum Match 100%
               Listing first 45 summaries

Database :   N_Geneseq_0601:*
       1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
       2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
       3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
       4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
       5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
       6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
       7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
       8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
       9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
      10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
      11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
      12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
      13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
      14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
      15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
      16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
      17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
      18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
      19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
      20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
      21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
      22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2481	100.0	2958	21	AAA58453	Human prostate can
2	2471.6	99.6	2478	21	AAA52810	Human sulphatase G
3	1782	71.8	2546	21	AAC76445	Human ORFX ORF2000
4	247.4	10.0	26664	21	AAA60207	Human prostate can
5	237	9.6	238	21	AAA80231	Human colon cancer
6	113.6	4.6	326	21	AAA60390	Murine prostate ca
7	90.2	3.6	1617	21	AAC34173	Arabidopsis thalia
8	90.2	3.6	2853	21	AAC49965	Arabidopsis thalia
9	58.4	2.4	2517	22	AAC89621	S. cerevisiae YKR0
10	51.4	2.1	1501	21	AAA12583	cDNA encoding a me
11	50.6	2.0	464	21	AAC38171	Arabidopsis thalia

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

c	23	48.2	1.9	16080	21	AAX15650	Nucleotide sequence
	24	45	1.8	795	19	AAV55830	FLGA insert stabil
	25	43.8	1.8	1468	21	AAV55496	Hela cell library
c	26	43.8	1.8	1528	20	AAX35555	Secreted protein c
c	27	43.6	1.8	217	16	AAQ97711	Branched-chain alp
	28	43.4	1.7	1567	19	AAX03145	Aspergillus terreu
	29	43	1.7	2125	20	AAX36279	Wheat Rht clone 5a
	30	42.8	1.7	453	20	AAX36263	DNA sequence obtai
c	31	42.8	1.7	2709	20	AAX36275	Composite DNA sequ
c	32	42.8	1.7	16442	20	AAX83006	Partial mouse WRN
	33	42.6	1.7	10732	18	AAA10594	Gene encoding a su
c	34	42.4	1.7	324	20	AAX36273	DNA sequence obtai
	35	42	1.7	2297	21	AAZ93309	Sequence encoding
c	36	42	1.7	38186	20	AAZ32028	Human METH1 relate
	37	41.4	1.7	38186	22	AAC90085	AC004449 cDNA clon
	38	41	1.7	422	21	AAC79295	Human lung tumour-
	39	41	1.7	2301	19	AAV20445	Human c-trk oncoge
	40	41	1.6	2301	21	AAA09303	Human TrK oncogene
	41	40.8	1.6	1429	21	AAZ93300	Sequence encoding
	42	40.4	1.6	15377	13	AAQ25975	MH mutant porcine
c	43	40.4	1.6	58857	21	AAA58471	Nucleotide sequenc
	44	40.2	1.6	1412	17	AAT26989	Thermus aquaticus
	45	39.8	1.6	6306	15	AAQ54841	Sequence encoding

ALIGNMENTS

RESULT   1
AAA58453

ID   AAA58453 standard; cDNA; 2958 BP.
XX
AC   AAA58453;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ss.
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   CDS   51..2531
FT   /*tag= a
FT   /product= "HPC2"
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US256055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.

```
XX
PT  Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT  antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS  Claim 3; Page 98-100; 157pp; English.
XX

CC  The present sequence is the coding sequence of the human prostate
CC  cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC  alleles of this gene cause a predisposition to cancer, particularly
CC  prostate cancer. This gene and its protein can be used in peptide and
CC  gene therapy for cancer patients, as well as being useful as diagnostic
CC  tools (both for cancer sufferers and those with a predisposition to the
CC  disease) and in the production of cancer drugs. This sequence was
CC  isolated by cloning and sequencing the region of the genome which
CC  appeared to cause a predisposition to prostate cancer.
XX
SQ  Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;
```

```
Query Match              100.0%;   Score 2481;   DB 21;   Length 2958;
Best Local Similarity    100.0%;
Matches 2481;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
                         Pred. No 0;
```

```
Qy     1 atgtgtggcgcttgctcgctctcgctgcgtcgcgtccgcggacgcaccagtcgcaggacgc 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    51 atgtgtggcgcttgctcgctctcgctgcgtcgcgtccgcggacgcaccagtcgcaggacgc 110

Qy    61 accatatcgcaggcaccaccgcccgcccgccccgccgacgcaccgcgtgcggcacctg 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   111 accatatcgcaggcaccaccgcccgcccgccccgccgacgcaccgcgtgcggcacctg 170

Qy   121 cgcacgcgagagaagccgcggaccgtcgggtgttcgggtgggcccaaacaccgtgtacctg 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 cgcacgcgagagaagccgcggaccgtcgggtgttcgggtgggcccaaacaccgtgtacctg 230

Qy   181 caggtcgttgcaggtcagcgggtagccgggactcggcggcgccgcctctcctctccgagttc 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   231 caggtcgttgcaggtcagcgggtagccgggactcggcggcgccgcctctcctctccgagttc 290

Qy   241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcagacacaagtta 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   291 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcagacacaagtta 350

Qy   301 aaggttgctccgcctggacaacatattcctgacacgatgcactcgtcgtctaatgttggggc 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   351 aaggttgctccgcctggacaacatattcctgacacgatgcactcgtcgtctaatgttggggc 410

Qy   361 ttaagtcggaatgatctcttaaagaaagaaaccggcgcttccaaagtgtgtactttctgga 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 ttaagtcggaatgatctcttaaagaaagaaaccggcgcttccaaagtgtgtactttctgga 470

Qy   421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 530

Qy   481 atagaactggctggcgtggggccccactctcgcccagaatacgagggatgaaacctgacagtt 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 atagaactggctggcgtggggccccactctcgcccagaatacgagggatgaaacctgacagtt 590

Qy   541 taccagatcccataacacaagtgaacagaggaggggaaagcaccacaccatggcagagtcca 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   591 taccagatcccataacacaagtgaacagaggaggggaaagcaccacaccatggcagagtcca 650

Qy   601 gaaaggcctcctcaggctcagcgactcttcagagtcctcccctgtc 660
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   651 gaaaggcctcctcaggctcagcgactcttcagagtcctcccctgtc 710

Qy   661 gagccacaccttccacatggttagccaggaggactccctggtc 720
         |||||||||||||||||||||||||||||||||||||||||||
Db   711 gagccacaccttccacatggttagccaggaggactccctggtc 770

Qy   721 gtagcttcatctgtaaagcttcacttaaagagaggaaacttcttggtgtgctcaaagcaaag 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   771 gtagcttcatctgtaaagcttcacttaaagagaggaaacttcttggtgtgctcaaagcaaag 830
```

```
Db   771 gtagcttcatctgtaaagcttcacttaaagagaggaaacttcttggtgtgctcaaagcaaag 830
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   781 gagatgggcctccccagtcggacagctgccatcgctccatcattgctgctgtctcaaggac 840

Db   831 gagatgggcctccccagtcggacagctgccatcgctccatcattgctgctgtctcaaggac 890
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   841 gggaaaagcaatcactcaatgaagagaagaggagattttggctgagaaagcttcattcca 900

Db   891 gggaaaagcaatcactcaatgaagagaagaggagattttggctgagaaagcttcattcca 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   901 gatcctggctgctgtctttcggtgtgtagaatgtcagagcttcattcaaccatc 960

Db   951 gatcctggctgctgtctttcggtgtgtagaatgtcagagcttcattcaaccatc 1010
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   961 tgtgaagaatgccacctttcagaggtaccaaaggaaaggcagatgccccgtggccttggtg 1020

Db  1011 tgtgaagaatgccacctttcagaggtaccaaaggaaaggcagatgccccgtggccttggtg 1070
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1021 gttcaacatggcccagcacactgctgcttgtgggacagcaagtccagtcagtggattggagag 1080

Db  1071 gttcaacatggcccagcacactgctgcttgtgggacagcaagtccagtcagtggattggagag 1130
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1081 tttgggccctgacacaccacttgctccttgtccgaatgaatgagacagctgctcagttcacaacctt 1140

Db  1131 tttgggccctgacacaccacttgctccttgtccgaatgaatgagacagctgctcagttcacaacctt 1190
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1141 cgcagccacaagattcaaaccacccagctcaacctcaatccaacctgctgcttccccctgctc 1200

Db  1191 cgcagccacaagattcaaaccacccagctcaacctcaatccaacctgctgcttccccctgctc 1250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1201 accagtttccgctgtaagaagaggggccccaccctcaaccagtccagtcccagtggccaatgg 1260

Db  1251 accagtttccgctgtaagaagaggggccccaccctcaaccagtccagtcccagtggccaatgg 1310
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1261 tgcctccctcaagtaccagctcccgctcctcacgtctctcctccagagttc 1320

Db  1311 tgcctccctcaagtaccagctcccgctcctcacgtctctcctccagagttc 1370
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1321 tgcaatcctgaggaattcattagttgaggcgctgcagctctcccaacttccagcagcgtg 1380

Db  1371 tgcaatcctgaggaattcattagttgaggcgctgcagctctcccaacttccagcagcgtg 1430
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1381 caggagtacaaggaggagtccgcaggaccagcgcccagcaggaagaaaggaagtcagtac 1440

Db  1431 caggagtacaaggaggagtccgcaggaccagcgcccagcaggaagaaaggaagtcagtac 1490
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1441 ccagaaatcatctccttggaacagggtctcacatcccatccagcaatcgaaatgtcagt 1500

Db  1491 ccagaaatcatctccttggaacagggtctcacatcccatccagcaatcgaaatgtcagt 1550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtgttgtggggcaca 1560

Db  1551 gccacacttgtcaacataagccccgacacgtctctgctactggactgtgttgtggggcaca 1610
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1561 tttgggcactgtccgtcattacggaagccgacactggccacctggct 1620

Db  1611 tttgggcagctgtccgtcattacggaagccgacactggccacctggct 1670
         ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1621 gctgtcgttgtgtccacccctgcacgcagcagatcacccagcaaggtcccaagtatctgctg 1680

Db  1671 gctgtcgttgtgtccacccctgcacgcagcagatcacccagcaaggtcccaagtatctgctg 1730
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1681 caggagagaacgcgcctggcctcacccttcaccctttgctggtgttggcc 1740

Db  1731 caggagagaacgcgcctggcctcacccttcaccctttgctggtgttggcc 1790
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1741 cccaccactcaaagcctgggctccagcagtaccagcagagtcctgagatccagaggtcac 1800

Db  1791 cccaccactcaaagcctgggctccagcagtaccagcagagtcctgagatccagaggtcac 1850
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1801 cacatcagtgatgattcctgccaaatgccttcaggaagggctgagatccagcctgca 1860

Db  1851 cacatcagtgatgattcctgccaaatgccttcaggaagggctgagatccagcctgca 1910
```

```
Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgattggaagaagagagttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1911 gtggaaagattgatcagttcgctgttgcgaacatgtgattggaagaagagagttcagacctgt 1970

Qy   1921 ctggtgcggcactgcaagacatgcgttgttgcgctggtgcacctctggctggaaa 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1971 ctggtgcggcactgcaagacatgcgttgttgcgctggtgcacctctggctggaaa 2030

Qy   1981 gtggtctattccgggggaacacatgccctgcgggctctgttctccggatgtgggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2031 gtggtctattccgggggaacacatgccctgcgggctctgttctccggatgtgggaaagatgcc 2090

Qy   2041 accccctgatacatgagacccctggaagaagagaagcagtggaaaag 2100
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2091 accccctgatacatgagacccctggaagaagagaagcagtggaaaag 2150

Qy   2101 acacacaagcacacacgtcccacaagccatcagcgtgtgggatgcggagtcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2151 acacacaagcacacacgtcccacaagccatcagcgtgtgggatgcggagtcatt 2210

Qy   2161 atgctgaaccacttcagcaacgctatgccaagtccccactccttcaagcccaaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2211 atgctgaaccacttcagcaacgctatgccaagtccccactccttcaagcccaaacttcagc 2270

Qy   2221 gagaaatgtggaagttgcctttgaccacacataaagtctgcttttggagacttccaacaatg 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2271 gagaaatgtggaagttgcctttgaccacacataaagtctgcttttggagacttccaacaatg 2330

Qy   2281 cccaagctgattccccactgaaagccctgtgtttgctggcgacatcgagagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2331 cccaagctgattccccactgaaagccctgtgtttgctggcgacatcgagagatggaggag 2390

Qy   2341 cgcaggagagaaagcgggagctgcggcgccctcctgtccaggggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2391 cgcaggagagaaagcgggagctgcggcgccctcctgtccaggggagctggca 2450

Qy   2401 ggcggcctgaggatgggggagcctcaagcagaagaggagccacaaggcc 2460
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2451 ggcggcctgaggatgggggagcctcaagcagaagaggagccacaaggcc 2510

Qy   2461 aagaaggtcagagcccagtga 2481
          |||||||||||||||||||||
Db   2511 aagaaggtcagagcccagtga 2531
```

```
RESULT  2
AAA52810
ID   AAA52810 standard; cDNA; 2478 BP.
XX
AC   AAA52810;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Human sulphatase G cDNA.
XX
KW   Human sulphatase G; hsG; chromosome 17p11.2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..2478
FT                   /*tag= a
FT                   /partial
FT                   /product= "hSG"
XX
PN   WO20003432 7-A1.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999;  99WO-AU01092.
XX
PR   09-DEC-1998;  98AU-0007624.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ, Litjens T,  Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   P-PSDB; AAY99850.
XX
PT   Novel isolated DNA sequence which encodes human sulphatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency -
XX
PS   Claim 2; Page 29-30; 33pp; English.
XX
CC   The present sequence encodes human sulphatase G (hSG). hSG is
CC   not a member of the well-characterised CTPSR sulphatase family.
CC   It belongs to a family showing sequence similarity to a sulphatase
CC   from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC   contains 23 exons and is located at chromosome 17p11.2. The present
CC   sequence is clone lambda29.1 from a human testes cDNA library. It was
CC   isolated using human EST sequences with sequence similarity to the
CC   non-CTPSR family as a probe to screen the library. The cDNA insert was
CC   subcloned and the DNA sequence of both strands was determined. The
CC   sequence may be used to treat a patient suffering from hSG deficiency
CC   by replacing, repairing, or compensating for a DNA sequence within that
CC   patient's genome.
XX
SQ   Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;

Query Match             99.6%;  Score 2471.6;  DB 21;  Length 2478;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 2474;  Conservative 0;  Mismatches   4;  Indels   0;  Gaps   0;
```

```
Qy      1 atgtggggcgctcttgtgctcgctgtcgcgtgcgctccggcgccgacgcgcagggagcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgtggggcgctcttgtgctcgctgtcgcgtgcgctccggcgccgacgcgcagggagcc 60

Qy     61 accatatcgcaggcgaaccccgccccgccgacccgagcccgcaagcgaaccgctgccgcctg 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 accatatcgcaggcgaaccccgccccgccgacccgagcccgcaagcgaaccgctgccgcctg 120

Qy    121 cgcacgcgagagaagcgcggacccgtcgggtgtccgggcggcccaaacacgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cgcacgcgagagaagcgcggacccgtcgggtgtccgggcggcccaaacacgtgtacctg 180

Qy    181 cagtgggtggcacgcggggtagccgggcactcggggcgggccgcgggcctcacacgtcctggga 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cagtgggtggcacgcggggtagccgggcactcggggcgggccgcgggcctcacacgtcctggga 240

Qy    241 aaccggtatctctcaactgtggagaagccgttcagaagactcatcgacgcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aaccggtatctctcaactgtggagaagccgttcagaagactcatcgacgcacaagtta 300

Qy    301 aaggttgctcgcctggacacacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 aaggttgctcgcctggacacacatattcctgacacgaatgcactggtctaatgttgggggc 360

Qy    361 ttaagtggaatgattctttaactttaaagagaaaacggggcttccaaagttgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ttaagtggaatgattctttaactttaaagagaaaacggggcttccaaagttgtactttctgga 420

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaaatatttctgttcattgaaatga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cctccacaactggaaaaatacctcgaagcaatcaaaaatatttctgttcattgaaatga 480

Qy    481 ataagaactggctgtcggcgcccccactctcgcccagaatacgaggatgaaaccatgacagt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ataagaactggctgtcggcgcccccactctcgcccagaatacgaggatgaaaccatgacagt 540

Qy    541 tacagatccccatacacagtgaacagagagggaaagacacaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tacagatccccatacacagtgaacagagagggaaagacacaccatggcagagtcca 600
```

```
Qy    601 gaaaggcctccaagcttcaagctccagtccagaagcgatcttcaagaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gaaaggcctccaagcttcaagctccagtccagaagcgatcttcaagaatgaaaat 660

Qy    661 gagccacacttccacatggttagccagaaggtcaggactcttccctggtc 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gagccacacttccacatggttagccagaaggtcaggactcttccctggtc 720

Qy    721 gtagcttcatctgtaagcttaacttcttggtgtctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gtagcttcatctgtaagcttaacttcttggtgtctcaaagcaaag 780

Qy    781 gggaaagcatcactcatgaagagaagagaagagatttt 840
          ||||||||||||||||||||||||||||||||||||||
Db    781 gagatgtgggcctccagtgggacagctgccatgctctccatcattgctgctcaaggac 840

Qy    841 gggaaagcatcactcatgaagagaagagaagagatttt 900
          |||||||||||||||||||||||||||||||||||||
Db    841 gggaaagcatcactcatgaagagaagagaagagatttt 900

Qy    901 gatcctggtcgtgcttttgtgttagaaatgcttcattcaaccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    901 gatcctggtcgtgcttttgtgttagaaatgcttcattcaaccatc 960

Qy    961 tgtgagaatgcaccttcagagggtaccaagagatgccccgtgggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    961 tgtgagaatgcaccttcagagggtaccaagagatgccccgtgggtg 1020

Qy   1021 gttcacatggcccaagcatctgtgcttgttgtgaccagcagtgg 1080
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1021 gttcacatggcccaagcatctgtgcttgttgtgaccagcagtgg 1080

Qy   1081 tttgggcctgaacacccagcacttggttcctgaatgagaactgcctcagttcacaacctt 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 tttgggcctgaacacccagcacttggttcctgaatgagaactgcctcagttcacaacctt 1140

Qy   1141 cgcagccacacagattcaaccatcatccaccgaacatctccccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 cgcagccacacagattcaaccatcatccaccgaacatctccccctgctc 1200

Qy   1201 accagttccgctgtaggaaggagggcccccacccccagtcgtgttcaggtgaa 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 accagttccgctgtaggaaggagggcccccacccccagtcgtgttcaggtgaa 1260

Qy   1261 tgcctcctccaagtaccagcctccgtccaagagggaggatggccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 tgcctcctccaagtaccagcctccgtccaagagggaggatggccattattact 1320

Qy   1321 tgcaatcctgaggaattcaatagttgaggcgcttcccaacttccagcagccgtg 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 tgcaatcctgaggaattcaatagttgaggcgcttcccaacttccagcagccgtg 1380

Qy   1381 caggagtacaggaggtcgcaggacgcccagccaccgcagagagaaaagaagtcagtac 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 caggagtacaggaggtcgcaggacgcccagccaccgcagagagaaaagaagtcagtac 1440

Qy   1441 ccagaaatcatctccttggaacaggtctgcaatccgatgaagaattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ccagaaatcatctccttggaacaggtctgcaatccgatgaagaattcgaaatgtcagt 1500

Qy   1501 gccacacttgtcaacataagccccgacaacgtctctgctactgactgtggttgaggggcaca 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gccacacttgtcaacataagccccgacaacgtctctgctactgactgtggttgaggggcacg 1560

Qy   1561 tttgggcagctggcgtgcattacggacaggtcctgggacccggt 1620
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1561 tttgggcagctggcgtgcattacggacaggtcctgggacccggt 1620

Qy   1621 gctgtgtttgtgtccacctgcacgcgagatcccaagtatctttgctg 1680
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 actgtgtttgtgtccacctgcacgcgagatcccaacgcacacacggcttcaacctgccaagtatctttgctg 1680
```

```
Qy   1681 caggagagaacgcgcctttgggcatctttgggaaaagcgcttcacccttttgctgttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 caggagagaacgcgcctttgggcatctttgggaaaagcgcttcacccttttgctgttgcc 1740

Qy   1741 cccaaccagtcaaagctccggctcttggcatcttcgggaagactaccaacacaacagtgcac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 cccaaccagtcaaagctccggctcttggcatcttcgggaagactaccaacacaacagtgcac 1800

Qy   1801 cacatcagtatgatgatcctgccaaatgcctttcaggactcaagatctctgca 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 cacatcagtatgatgatcctgccaaatgcctttcaggactcaagatctctgca 1860

Qy   1861 gtggaaagattgatcagttccctgttgcgaacatgtgaagttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 gtggaaagattgatcagttccctgttgcgaacatgtgaagttcagacctgt 1920

Qy   1921 ctggtgcgcggcactgcaagagctgcctgtttgggctgttggctgtgcacctctggctgaaa 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ctggtgcgcggcactgcaagagctgcctgtttgggctgttggctgtgcacctctggctgaaa 1980

Qy   1981 gtggtctattccgggacacatgcccctgcgaggctcggatgggagagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 gtggtctattccgggacacatgcccctgcgaggctcggatgggagagatgcc 2040

Qy   2041 accctccttatacatgaagcacaccctggaagagtcgagtcgtcttcgagatg 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 accctccttatacatgaagcacaccctggaagagtcgagtcgtcttcgagatg 2100

Qy   2101 acacacacagacacgtcccaagccatcagcgtgggatgtgaacgcggagtcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 acacacacagacacgtcccaagccatcagcgtgggatgtgaacgcggagtcatt 2160

Qy   2161 atgctgaaccactcagccagctatgccaagtgccctctcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 atgctgaaccactcagccagctatgccaagtgccctctcagccccaacttcagc 2220

Qy   2221 gagaaagtgggagttgcctttgaccacacagaagtctgttcttgagacttttccacaatg 2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 gagaaagtgggagttgcctttgaccacacagaagtctgttcttgagacttttccacaatg 2280

Qy   2281 cccaagctgattccccactgaaagcccctgtttgctgtgcggcaatggagatgagggag 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 cccaagctgattccccactgaaagcccctgtttgctgtgcggcaatggagatgagggag 2340

Qy   2341 cgcaggagaagacgggagactgcggcggcgccctctcaggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 cgcaggagaagacgggagactgcggcggcgccctctcaggagctggca 2400

Qy   2401 ggcggcctggaggatggggagcctcaagcagcagaggagccaaggcc 2460
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ggcggcctggaggatggggagcctcaagcagcagaggagccaaggcc 2460

Qy   2461 aagaaaggtcagagccag 2478
          ||||||||||||||||||
Db   2461 aagaaaggtcagagccag 2478
```

```
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;    99US-0127607.
PR   02-APR-1999;    99US-0127636.
PR   05-APR-1999;    99US-0127728.
PR   30-MAR-2000;  2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42236.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3179-3180; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; immunosuppressive;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   anticonvulsant; antiarthritic; immunosuppressant;
CC   osteopathic;antihypertensive; cardiant; thrombolytic; coagulant; vasotropic;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;
```

Query Match   71.8%;   Score 1782;   DB 21;   Length 2546;
Best Local Similarity   99.1%;   Pred. No. 0;
Matches 1813;   Conservative   0; Mismatches   15;   Indels   2;   Gaps   2;

```
Qy      654 tgaaaatgagccacacttccacatggttcagcagagagagggtcagggacttc 713
            |||.||||  |||  |||||||||||||||||||||||||||||||||||||||
Db      371 aagcaaaggagaatgggcctcccagttgggacagctgccatcgctcccatcattgctgctg 430

Qy      654 tgaaaatgagccacacttccacatggttcagcagagagagggtcagggacttc 713
            || |  |||  |  | |||||||||||||||||||||||||||||||||||||
Db      251 tgtataatcaagtcgtgtttcctccagtgttagccagagaaaggtcagggactcttc 310

Qy      714 cctggtcgttagcttcattctgtaagcttcacttaaagagagagaaactt-cctggtgctca 772
            ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      311 cctggtcgtagcttccattcctgtaagcttcacttaaagagagagaaacttactctggtgctca 370

Qy      773 aagcaaaggagaatgggcctcccagttgggacagctgccatcgctcccatcattgctgctg 832
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      431 tcaaggacgggaaaagcatcactcaatcatcgaaggaagaagagagagattttggctgaagagcgtgta 490

Qy      833 tcaaggacgggaaaagcatcactcaatcatcgaaggaagaagagagagattttggctgaagagctgcta 892
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      491 ctcctccagatcctggtgctgcttttgtggtggttggtagaatgtccagatgaaaagcttcatcc 550

Qy      893 ctcctccagatcctggtgctgcttttgtggtggttggtagaatgtccagatgaaaagcttcatcc 952
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      551 aacccatctgtgaaggatgccacacctttcagaggttaccaagggaaaggcaagatgccccgg 610

Qy      953 aacccatctgtgaaggatgccacacctttcagaggttaccaagggaaaggcaagatgccccgg 1012
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      611 cccttggttggttcacatggccccagccatctgttgctgttggacagagtaccagcagtgga 670

Qy     1013 cccttggttggttcacatggccccagccatctgttgctgttggacagagtaccagcagtgga 1072
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      671 tggaagaggtttgggcctgacaccccagccactcggtcctgaatgagaactgtgtcctcagttc 730

Qy     1073 tggaagaggtttgggcctgacaccccagccactcggtcctgaatgagaactgtgtcctcagttc 1132
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      731 acaacccttcgcagccacaagattcaaacccagtctcaaacctcatccacactcggacactctcc 790

Qy     1133 acaaccctccgcagccacaagatccaaacccagtctcaaacctcatccacactcggacatctctcc 1192
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      791 cctcgctcaccagttccgctgtaagaagaaggagcccccaccccttcagttgctgccatggttc 850

Qy     1193 cctcgctcaccagttccgctgtaagaagaaggagcccccaccccttcagttgctgccatggttc 1252
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      851 agggtgaatgcttcctccaagtaccagctccgtccaagcgaggggagtggagaggggatgcca 910

Qy     1253 agggtgaatgcttcctccaagtaccagctccgtccaagcgaggggagtggagaggggatgcca 1312
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      911 ttattacttgcaatcctgaggaattcataagttgaggccgtgcagcttccaactccagc 970

Qy     1313 ttattacttgcaatcctgaggaattcataagttgaggccgtgcagcttccaactccagc 1372
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      971 agagcgtgcaggagtacaggacggccagcggcccagcccagcagagagaaga 1030

Qy     1373 agagcgtgcaggagtacaggacggccagcggcccagcccagcagagagaaga 1432
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1031 gtcagtacccagaaatcatcttcctttggaacaggtctcgcatccccgatgaagatcgaa 1090

Qy     1433 gtcagtacccagaaatcatcttcctttggaacaggtctcgcatccccgatgaagatcgaa 1492
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1091 atgtcaagtgccacacttgtcaaacataagccccgacacagtctctgctactggtggtcg 1150

Qy     1493 atgtcaagtgccacacttgtcaaacataagccccgacacagtctctgctactggtggtcg 1552
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1151 agggcacattgggcagctgccgtcattacggaggacaggtgacaggtcctggca 1210

Qy     1553 agggcacattgggcagctgtgtccgtcattacggagacaccaggtggacaggtcctggca 1612
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1211 cctggctggctgcttgttctgtccaactgcacgcagatcaccacacccgggcttgccaagta 1270

Qy     1613 cctggctggctgcttgttctgtccaactgcacgcagatcaccacacccgggcttgccaagta 1672
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1271 tctctggctgcagagaacgcgcctgccatcttggacttcttgggaatccag 1330

Qy     1673 tctctggctgcagagaacgcgcctgccatcttgggaaatccgcttcaccctttgctg 1732
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1331 tggttgcccccaacacagcctcccaacagtaccacccaaccacacagtgccaaggag 1390

Qy     1733 tggttgcccccaacacagcctcccaacagtaccacccaaccacacagtgccaggag 1792
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1391 tcctgccacacatcaagtatgatcagttcgctgttgcgaacacatgtgattgggaagagttc 1450

Qy     1793 tcctgccacacatcaagtatgatcagttcgctgttgcgaacacatgtgattgggaagagttc 1852
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1391 tcctgccacacatcaagtatgatcagttcgctgttgcgaacacatgtgattgggaagagttc 1450

Qy     1853 gtcctgcagtgaaaagtctcgttgttgcgaacacatgtgattgggaagagttc 1912
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1451 gtcctgcgtggaaagattgatcagttcgctgttgcgaacatgtgattggaagagttc 1510
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1913 agacctgctcgtggcgcgactgcaagcacatgcgtttgctgtcgcgctggtcacacctctg 1972
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 agacctgctcgtggcgcgactgcaagcacatgcgtttggctgtcgcgctggtgcacacctctg 1570
Qy   1973 gctggaagtgtctattccgggggaccaccatgccctccgggactctcgtctccggatggga 2032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 gctggaagtggtctattccgggggaccaccatgccctccgggactctcgtctccggatggga 1630
Qy   2033 aagatgccacccctcgatacatgaagccaccctggaagatggtttggaagaggaagcag 2092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1631 aagatgccacctcctcgatacatgaagccaccctggaagatggtttggaagaggaagcag 1690
Qy   2093 tggaaaagaacaacagcacacagtcccaagctgcggatgcggatcggaggaacgcg 2152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1691 tggaaaagaacaacagcacacagtcccaagctgtgggatcggaggatcggaggaacgcg 1750
Qy   2153 agttcattatgtctgaaccacttcagccaagctgtatgccagccttcagccgtgtccccca 2212
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1751 agttcattatgtctgaaccacttcagccaagctgtatgccagccttcagccgtgtccccca 1810
Qy   2213 acttcacgcagaaagtgggagtgttgcctttgaccacacagaagtctgcttggagacttc 2272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1811 acttcacgcagaaagtgggagtgttgcctttgaccacacagaagtctgcttggagacttc 1870
Qy   2273 caacaatgcccaagctgatt-cccccactgaaagccctgtttgctgtcgcgacatcgaggag 2331
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1871 caacaatgcccaagctgattcccccactgaaagccctgtttgctgtcgcgacatcgaggag 1930
Qy   2332 atggagagcgcaggggaggaagactgcggggtcgggtcgcggtcgcggtcgcggtcccagg 2391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1931 atggagagcgcaggggaggaagactgcggggtcgggtcgcggtcgcggtcgcggtcccagg 1990
Qy   2392 gagctgcggcaggccggcctggaggatgggggagctcagcagaacggggcccacacagaggag 2451
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1991 gagctgcggcaggccggcctggaggatgggggagctcagcagaacggggcccacacagaggag 2050
Qy   2452 ccacaggccaagaagtcaagagccagtga 2481
          |||||||||||||||||||||||||||||
Db   2051 ccacaggccaagaagtcaagagccagtga 2080
```

```
RESULT   4
AAA60207

ID   AAA60207 standard; DNA; 26664 BP.
XX
AC   AAA60207;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 genomic sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             910..26039
FT                   /*tag= a
FT                   /product= "HPC2"
FT                   /note= "this sequence contains introns"
FT                   /transl_except= (pos:23892..23895,aa:Glu)
FT   exon            910..1154
FT                   /*tag= b
FT                   /number= 1
FT   exon            1736..1786
FT                   /*tag= c
FT                   /number= 2
FT   exon            1925..1995
FT                   /*tag= d
FT                   /number= 3
FT   exon            3025..3089
FT                   /*tag= e
FT                   /number= 4
FT   exon            4361..4418
FT                   /*tag= f
FT                   /number= 5
FT   exon            5582..5650
FT                   /*tag= g
FT                   /number= 6
FT   exon            7075..7194
FT                   /*tag= h
FT                   /number= 7
FT   exon            8186..8244
FT                   /*tag= i
FT                   /number= 8
FT   exon            12878..12936
FT                   /*tag= j
FT                   /number= 9
FT   exon            13032..13104
FT                   /*tag= k
FT                   /number= 10
FT   exon            13756..13868
FT                   /*tag= l
FT                   /number= 11
FT   exon            15283..15378
FT                   /*tag= m
FT                   /number= 12
FT   exon            16278..16416
FT                   /*tag= n
FT                   /number= 13
FT   exon            16498..16583
FT                   /*tag= o
FT                   /number= 14
FT   exon            18583..18701
FT                   /*tag= p
FT                   /number= 15
FT   exon            20349..20445
FT                   /*tag= q
FT                   /number= 16
FT   exon            22879..22917
FT                   /*tag= r
FT                   /number= 17
FT   exon            22172..22310
FT                   /*tag= s
FT                   /number= 18
FT   exon            23045..23154
FT                   /*tag= t
FT                   /number= 19
FT   exon            23795..23895
FT                   /*tag= u
FT                   /number= 20
FT   exon            23973..24093
FT                   /*tag= v
FT                   /number= 21
FT   exon            24354..24432
FT                   /*tag= w
FT                   /number= 22
FT   exon            25026..25170
FT                   /*tag= x
FT                   /number= 23
FT   exon            25812..26036
FT                   /*tag= y
FT                   /number= 24
FT   polyA_signal    26447..26452
FT                   /*tag= z
XX
PN   WO20002027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
PR   06-NOV-1998;  98US-0107468.
XX
```

```
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Claim 3; Page 108-122; 157pp; English.
XX
CC   The present sequence is the genomic sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;

Query Match         10.0%;  Score 247.4; DB 21; Length 26664;
Best Local Similarity   97.7%;  Pred. No. 7.8e-53;
Matches  251; Conservative   0; Mismatches   6;  Indels   0; Gaps   0;

Qy       1 atgtggggccgctttgctcgctgctgcggtcgccgggccggacgcacgacaatgtcgcaggggacgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     910 atgtggggccgctttgctcgctgctgcggtcgccgggccggacgcacgacaatgtcgcaggggacgc 969

Qy      61 accatatcgcaggcaaccgaccccgcccgccgccgaggcgccggacgcaagtcggtcgcggcaacctg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     970 accatatcgcaggcaaccgaccccgcccgccgccgaggcgccggacgcaagtcggtcgcggcaacctg 1029

Qy     121 cgcacgcgagagaaagcgcggacgacgtcggggtcgctccggcggcccaaacaaccggtgtacctg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1030 cgcacgcgagagaaagcgcggacgacgtcggggtcgctccggcggcccaaacaaccggtgtacctg 1089

Qy     181 cagggtggttggtgcagcgggtagccgggactcgggtcgccgcgctctacgtctctccgagttc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1090 cagggtggttggtgcagcgggtagccgggactcgggtcgccgcgctctacgtctctccgagttc 1149

Qy     241 aaccggtatctcttcaa 257
           ||||||| || ||
Db    1150 aaccggtcagtcaacga 1166
```

```
XX
PA   (FARB ) BAYER CORP.
XX
PI   Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI   Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-087220/07.
XX
PR   10-JUN-1998;  98US-0088801.
XX
PS   Claim 15; Page 258; 469pp; English.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer  -
XX
CC   AA279917 to AA280766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which
CC   can be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g.
CC   carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC   can be used for developing agents for the diagnosis and treatment of
CC   disorders involving unwanted cell proliferation, such as neoplasia,
CC   dysplasia or hyperplasia.
XX
SQ   Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;

Query Match          9.6%;  Score 237; DB 21; Length 238;
Best Local Similarity  100.0%;  Pred. No. 4.4e-51;
Matches  237; Conservative   0; Mismatches   0;  Indels   0; Gaps   0;

Qy     176 acctgcaggtggtggcagcgggttagccgggccgccggacttcacgtctctaagtctctccg 235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 acctgcaggtggtggcagcgggttagccgggccgccggacttcacgtctctaagtctctccg 60

Qy     236 agttcaaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcaca 295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 agttcaaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcaca 120

Qy     296 agttaaagtttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttg 355
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 agttaaagtttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttg 180

Qy     356 ggggcttaagtggaatgattcttacttttaaagaaaccgggcttccaaagtgtgtac 412
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ggggcttaagtggaatgattcttacttttaaagaaaccgggcttccaaagtgtgtac 237
```

```
RESULT   5
ID   AA280231 standard; cDNA; 238 BP.
XX
AC   AA280231;
XX
DT   07-APR-2000  (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW   hyperplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO9964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999;  99WO-IB01062.
```

```
RESULT   6
ID   AAA60390 standard; cDNA; 326 BP.
XX
AC   AAA60390;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Murine prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Mouse; prostate cancer predisposing gene; HPC2; gene therapy;
KW   peptide therapy; drug design; ss.
XX
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            51..269
FT                  /*tag= a
FT                  /product= "Mm.HPC2"
FT                  /partial
FT                  /note= "this sequence contains no termination codon"
```

```
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;                                                          99WO-US26055.
XX
PR   06-NOV-1998;                                                          98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI: 2000-376481/32.
DR   P-PSDB; AAB07230.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Disclosure; Page 151; 157pp; English.
XX
CC   The present sequence is the coding sequence of the murine prostate
CC   cancer predisposing gene Mn. HPC2, the human homologue of which is found
CC   on chromosome 17p. Some alleles of this gene cause a predisposition to
CC   cancer, particularly prostate cancer. This gene and its protein can be
CC   used in peptide and gene therapy for cancer patients, as well as being
CC   useful as diagnostic tools (both for cancer sufferers and those with a
CC   predisposition to the disease) and in the production of cancer drugs.
XX
SQ   Sequence 326 BP; 39 A; 104 C; 127 G; 56 T; 0 other;

Query Match        4.68;  Score 113.6; DB 21;  Length 326;
Best Local Similarity  78.9%;
Matches 150;  Conservative  0;  Mismatches  34;  Indels  6;  Gaps  1;

Qy     58 cgcacacatatcgcaggcaccccgccccgcgaggcggcgcaaggaggaccccggtcgcggcac  117
          |||||||| ||||||||| | || || |||||||||||||||||||||||||||| ||
Db     90 cgcaccatgtccgcagggttcggctcgtccggccggccaccaaggaccacctccgacac  149

Qy    118 ctgcgcacgcgagagaagcgcggacaccgtcgggtgctccggcgcccaaacaccgtgtac  177
          ||||| ||||||||||||||||||||| || ||| |||||||||||||||||||||
Db    150 ctgcgtacgcgggagagaagcgcggcccc------gggtcccggggccgaacaccgtgtac  203

Qy    178 ctgcgcaggtggtggcagcggttagccggggactcgggcgccgcgctctacgtcttcctccgag  237
          |||||||||| |||||||||| ||||||||| |||||||||||| |||||||| ||
Db    204 ctgcgcaggtgttggcggcgggcgcgggaccgcgggggctgctctctcatgtcttctccgaa  263

Qy    238 ttcaaccggt  247
          | ||| |||
Db    264 tacaacaggt  273
```

```
RESULT    7
AAC34173
ID   AAC34173 standard; DNA; 1617 BP.
XX
AC   AAC34173;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 5716.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP103405-A2.
XX
PD   06-SEP-2000.
XX
```

PF	25-FEB-2000;	2000EP-0301439.
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	30-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0137082.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142290.
PR	08-JUL-1999;	99US-0142803.

PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   26-JUL-1999;   99US-0145224.
PR   27-JUL-1999;   99US-0145276.
PR   28-JUL-1999;   99US-0145513.
PR   27-JUL-1999;   99US-0145518.
PR   22-JUL-1999;   99US-0145919.
PR   02-AUG-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   03-AUG-1999;   99US-0146389.
PR   04-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   05-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149990.
PR   25-AUG-1999;   99US-0150556.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   20-SEP-1999;   99US-0155119.
PR   22-SEP-1999;   99US-0155179.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.

PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   14-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   21-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   25-OCT-1999;   99US-0160815.
PR   25-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   26-OCT-1999;   99US-0161404.
PR   26-OCT-1999;   99US-0161405.
PR   26-OCT-1999;   99US-0161406.
PR   25-OCT-1999;   99US-0161359.
PR   25-OCT-1999;   99US-0161360.
PR   25-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   3.6%;   Score 90.2;   DB 21;   Length 1617;
Best Local Similarity   53.9%;   Pred. No. 3.1e-13;
Matches 185;   Conservative   0;   Mismatches 158;   Indels   0;   Gaps   0;

Qy   1970 ctggctgaaagttgttcattccgggacaccatgccctgcgaggctctggtccggatgg 2029
                |||| |||| ||| || | ||||||| | | | |||| ||| | |||| | |||
Db   1093 cggggtgaaaatggtttactcgggagacacaggcctgtgtcctgagatggttgaagctt 1152

Qy   2030 ggaaagatgccaccctcctgatacatgaagcacacctggaagatggttgttggaagaaag 2089
                |||| | | ||||| | | | | || || || ||| | | ||| || || |||
Db   1153 caaaaggagcaacagttcttatacatgaggcgacgttgaggatgcgttgattgaaagaag 1212

Qy   2090 cagtggaaaagaacacacagcacaacgtcccaagccatcagcgtgggggatcgggatgaacg 2149
                | | ||||| | | |||||||| ||| ||||| || ||| |||| |||| ||||| |
Db   1213 ctgtggctaaaaaccatagcacaactaaagaagcaatccaaagttggatcttcagcaggtg 1272

Qy   2150 cggagttcattatgctgaaccacactcagccagcgcctatgccaagtccccctcttcagcc 2209
                |||| | ||| || || | | ||||| | ||||||| | | | |||||||| | |||||
Db   1273 tataccgaaccgtactgacacatttagccagagatacccaaagatcccggtgattgacg 1332

Qy   2210 ccaacttcagcgagaagtggagttgccttgaccacacatgaaggtctgcttggagact 2269
                || ||||| |||| ||| | | | || ||| ||| ||||||||| | || | |||
Db   1333 aatcacatatgcacaatacatgcattgcctttgatgatgagtataaacatgtgcggatt 1392

Qy   2270 ttccaaacaatgcccaagtgattcccccactgaaagcccctgtt 2312
                | | ||||| || | ||||| | ||| ||| | || | | |
Db   1393 tacatgtgcttccaaagatttacttcacttcaaaaactctatt 1435

RESULT   8
AC49965
ID   AAC49965   standard;   DNA;   2853 BP.
XX
AC   AAC49965;
XX
DT   18-OCT-2000   (first entry)

```
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 63099.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   05-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   06-MAY-1999; 99US-0132863.
PR   07-MAY-1999; 99US-0134256.
PR   11-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   14-MAY-1999; 99US-0134768.
PR   18-MAY-1999; 99US-0134941.
PR   19-MAY-1999; 99US-0135124.
PR   20-MAY-1999; 99US-0135353.
PR   21-MAY-1999; 99US-0135629.
PR   24-MAY-1999; 99US-0136021.
PR   25-MAY-1999; 99US-0136392.
PR   27-MAY-1999; 99US-0136782.
PR   28-MAY-1999; 99US-0137222.
PR   01-JUN-1999; 99US-0137528.
PR   03-JUN-1999; 99US-0137502.
PR   04-JUN-1999; 99US-0137724.
PR   07-JUN-1999; 99US-0138094.
PR   08-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   10-JUN-1999; 99US-0139119.
PR   14-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   16-JUN-1999; 99US-0139492.
PR   17-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   15-JUL-1999; 99US-0140005.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
PR   20-JUL-1999; 99US-0144884.
PR   21-JUL-1999; 99US-0144814.
PR   21-JUL-1999; 99US-0145086.
PR   21-JUL-1999; 99US-0145087.
PR   22-JUL-1999; 99US-0145088.
PR   22-JUL-1999; 99US-0145085.
PR   22-JUL-1999; 99US-0145089.
PR   22-JUL-1999; 99US-0145192.
PR   23-JUL-1999; 99US-0145145.
PR   23-JUL-1999; 99US-0145218.
PR   23-JUL-1999; 99US-0145224.
PR   26-JUL-1999; 99US-0142276.
PR   27-JUL-1999; 99US-0145913.
PR   27-JUL-1999; 99US-0145918.
PR   28-JUL-1999; 99US-0145919.
PR   02-AUG-1999; 99US-0145951.
PR   02-AUG-1999; 99US-0146386.
PR   02-AUG-1999; 99US-0146388.
PR   03-AUG-1999; 99US-0146389.
PR   04-AUG-1999; 99US-0147038.
PR   04-AUG-1999; 99US-0147204.
PR   05-AUG-1999; 99US-0147192.
PR   05-AUG-1999; 99US-0147260.
PR   06-AUG-1999; 99US-0147302.
PR   06-AUG-1999; 99US-0147303.
PR   06-AUG-1999; 99US-0147416.
PR   09-AUG-1999; 99US-0147493.
PR   09-AUG-1999; 99US-0147935.
PR   10-AUG-1999; 99US-0148171.
PR   11-AUG-1999; 99US-0148319.
PR   12-AUG-1999; 99US-0148341.
PR   13-AUG-1999; 99US-0148565.
PR   13-AUG-1999; 99US-0148684.
PR   16-AUG-1999; 99US-0148968.
PR   17-AUG-1999; 99US-0149175.
PR   18-AUG-1999; 99US-0149426.
PR   20-AUG-1999; 99US-0149722.
PR   20-AUG-1999; 99US-0149723.
PR   20-AUG-1999; 99US-0149902.
PR   23-AUG-1999; 99US-0149929.
PR   23-AUG-1999; 99US-0149930.
PR   25-AUG-1999; 99US-0150566.
PR   26-AUG-1999; 99US-0150884.
PR   27-AUG-1999; 99US-0151065.
PR   27-AUG-1999; 99US-0151066.
```

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151203.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0155458.
PR 29-SEP-1999; 99US-0155596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0180029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0160981.
PR 14-OCT-1999; 99US-0161089.
PR 18-OCT-1999; 99US-0159931.
PR 21-OCT-1999; 99US-0156637.
PR 25-OCT-1999; 99US-0155984.
PR 25-OCT-1999; 99US-0159638.
PR 25-OCT-1999; 99US-0160741.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 26-OCT-1999; 99US-0160770.
PR 26-OCT-1999; 99US-0160714.
PR 26-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   53.9%;   Score 90.2; DB 21; Length 2853;
Best Local Similarity   53.9%;
Matches 185; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 1970 ctgtgctgaaagtggtctcattccgggacacatgccctgcgaggctctggtccggatgg 2029
          |||| |||||||| | ||||||||| |||||||| ||||||| ||||||| ||||||
Db 2333 cggggtggaaagtgttactcgggagacacaagcgtttcctgagatcgttgaagctt 2392

Qy 2030 ggaaagatgccacctcctgatacatgaagccacccctgggaagatggtttggaagaggaag 2089
          | ||| |||||||| | |||| || |||||||||| |||||| | ||| |||||||||
Db 2393 caaaaggcaacagtcttatacatgaggcgacgtttgaagatgcgttcggtaagaag 2452

Qy 2090 cagtggaaagacaacacagcaccaacgtccaagccatcagcgtggggatgcgggatgc 2149
          | || ||| | |||| | |||| | ||||| |||||| ||| | ||| |||
Db 2453 ctgtggctaaaacatagcacaactaaagaaagctggatcttcagcaggtg 2512

Qy 2150 cggagttcattatgctgaacacttcagccagcgctatgccaagtccccctcttcagcc 2209
          | || ||| || |||||||||||||||| || |||||||||| || | || ||||
Db 2513 tataccgacctgactgacacatttagccagagaaaacactgcactattaatcggtg 2572

Qy 2210 ccaacttcagcgagaaaggtgggaggttgttggcacactagcacgcctatgccaagtccc 2269
          | ||||||| | |||||| || ||| ||||| |||| || |||| ||||| |||||||
Db 2573 aatcacacatgcacaaatacatgcattgcttgatatgatgagtataaacatggcggatt 2632

Qy 2270 tccoaacaatgcccaagctgattccccactgaaagccctgtt 2312
          ||| ||| ||| | || ||| ||| |||| ||| |||||
Db 2633 tacatgtgctccaaagatttacttacttcaaaactctatt 2675

RESULT 9

AAC89621

ID AAC89621 standard; DNA; 2517 BP.
XX
AC AAC89621;
XX
DT 08-MAR-2001 (first entry)
XX
DE S. cerevisiae YKR079C gene.
XX
KW Yeast; germination; proliferation; essential gene; antifungal agent;
KW insecticide; herbicide; anti-proliferation drug; cancer; psoriasis;
KW restenosis; YKR081C; YFR003C; YGR277C; YGR278W; YKR071C; YKR079C;
KW YKR083C; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200071161-A1.
XX
PD 30-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13017.
XX
PR 21-MAY-1999; 99US-0315794.
PR 02-SEP-1999; 99US-0389341.
XX
PT Identifying antifungal compounds which target yeast essential genes
PT comprises use of novel Saccharomyces cerevisiae essential genes
PT YFR003C, YGR277C, YGR278W, YKR071C, YKR079C or YKR081C -
XX
PS Example 5; Fig 29; 127pp; English.
XX
CC The present invention provides methods of identifying antifungal agents
CC using the coding and protein sequences of several yeast genes. These are
CC essential for the germination and proliferation of Saccharomyces
CC cerevisiae, and include YKR081C, YFR003C, YGR277C, YGR278W, YKR071C,
CC YKR079C and YKR083C. The sequences can also be used to identify compounds
CC for use as herbicides, insecticides and anti-proliferation drugs which
CC can be used in the treatment of cancer, psoriasis and restenosis. This is
CC because they can be used to identify plant, insect and human homologues
CC of the yeast genes.
XX
SQ Sequence 2517 BP; 881 A; 417 C; 496 G; 723 T; 0 other;
XX
DR WPI; 2001-025092/03.
DR P-PSDB; AAB49964.
XX
PI Roberts CJ;
XX
PA (ROSE-) ROSETTA INPHARMATICS INC.

Query Match   51.2%;   Score 58.4; DB 22; Length 2517;
Best Local Similarity   51.2%;
Matches 168; Conservative 0; Mismatches 151; Indels 9; Gaps 1;

Qy 2021 tccggatggggaaagatgccacccctgatacatgaagccacccctggaagatggttcgg 2080
          ||| ||| |||| || | |||| |||||| | ||||||| |||| |||||| |||
Db 2126 tcgaaataggctatataatcagatctattaattcaacgaagctacactagaaaactcac 2185

Qy 2081 aagaggaagcagtggaaagacaacacagcacaacgtccaagccatcagcgtggggatgc 2140
          ||||| ||| | || ||| | |||| | || || | ||||| | | ||| | |||
Db 2186 tggaggatgcggctgaagaaaaacactgcactattaatgaagcaatcggtgttccgaca 2245

Qy 2141 ggatgaacgcggagttcattatgctgaacacttcagccagcgctatgccaagtcccc 2200
          | ||||| | | |||||| || || |||| | ||||||| |||||||| ||||||
Db 2246 aaatgaatgctagaagagttgatcttaacacacttcccagagatatcccaaattgccc 2305

```
Qy   2201 tcttcagcccaacttcagcagagaaagtgggag----------ttgcctttgaccacatga 2251
          |||||||||||||||||||||||||||||||||        ||||||||||||||||||||
Db   2306 aattagacaataatatgatgtgatgcgagagaatttgtctttgcttccgacagtatga 2365

Qy   2252 aggtctgtcttgggagacttccaacaatgcccaagctgattcccactgaaagccctgt 2311
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   2366 tcgttgattatgagaaaattggtgaacagcgcgtattttccactgctgaataaggcat 2425

Qy   2312 ttgctgccgaaatcgggagagtgggga 2339
          ||||||||| ||||| ||||| ||| ||
Db   2426 ttgttgaagaaaggaagaagagaaga 2453
```

```
RESULT  10
ID   AAA12583 standard; cDNA; 1501 BP.
XX
AC   AAA12583;
XX
DT   25-JUL-2000  (first entry)
XX
DE   cDNA encoding a membrane associated organizational protein (HJNCT).
XX
KW   Human; membrane associated organizational protein; HJNCT;
KW   cell proliferative disorder; cancer; autoimmune disorder;
KW   inflammatory disorder; neurological disorder; developmental disorder;
KW   vesicle trafficking; reproductive disorder; gastrointestinal disorder;
KW   renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;
KW   Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;
KW   irritable bowel syndrome; allergy; ss.
XX
OS   Homo sapiens.
XX
FH
FT   Key                 Location/Qualifiers
FT   CDS                 183..1268
FT                       /tag= a
FT                       /product= "membrane associated organizational protein"
XX
PN   WO200018915-A2.
XX
PD   06-APR-2000.
XX
PF   23-SEP-1999;  99WO-US220082.
XX
PR   25-SEP-1998;  98US-0155215.
PR   13-OCT-1998;  98US-0155251.
PR   04-MAY-1999;  99US-0172228.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yue H,  Lal P,  Corley NC,  Guegler KJ,  Baughn MR,  Lu AD,  Tang YT;
XX
DR   WPI; 2000-293154/25.
DR   P-PSDB; AAY04607.
XX
PT   The present sequence encodes a human membrane associated organizational
PT   protein (HJNCT). HJNCT is used for the diagnosis, treatment and
PT   prevention of cell proliferative disorders including cancer and
PT   autoimmune/inflammatory, neurological, developmental, vesicle
PT   trafficking, reproductive, gastrointestinal and renal disorders. These
PT   disorders may include atherosclerosis, leukaemia, allergies, rheumatoid
PT   arthritis, Alzheimer's disease, anxiety, diabetes, ovulatory defects,
PT   renal failure and irritable bowel syndrome. A vector expressing HJNCT,
PT   and an agonist of HJNCT can be used to treat or prevent a disorder
PT   associated with decreased expression or activity of HJNCT. An antagonist
```

```
CC   of HJNCT or a vector expressing the complement of a polynucleotide
CC   encoding HJNCT can be used to treat or prevent a disorder associated
CC   with increased expression or activity of HJNCT. Antibodies which bind
CC   HJNCT can be used for diagnosis of disorders associated with HJNCT
CC   expression or to monitor patients being treated with HJNCT, agonists,
CC   antagonists or inhibitors of HJNCT. Assays are preferably carried out on
CC   body fluids from a patient using radioimmunoassay, enzyme linked
CC   immunosorbent assays or fluorescent activated cell sorting assays.
CC   Polynucleotides encoding HJNCT are also used in hybridisation assays to
CC   determine absence, presence or excess expression of HJNCT and to monitor
CC   regulation of HJNCT levels during disease therapy.
XX
SQ   Sequence 1501 BP; 300 A; 443 C; 520 G; 238 T; 0 other;
```

```
Query Match           2.1%;  Score 51.4;  DB 21;  Length 1501;
Best Local Similarity  56.1%;  Pred. No. 0.0024;
Matches   97;  Conservative  0;  Mismatches  76;  Indels  0;  Gaps  0;
```

```
Qy   2296 ccactgaaagccctgtttgctgccgaaatcgggagagatgaggagccgcaggaggaagccg 2355
          ||| || ||||||||| |||| |||| ||||||| |||||||||
Db   1023 ccccgcgaggccctggagctgccgagccgtgaggagactcagcagcagaaggtgcaagg 1082

Qy   2356 gagctgcggcaggtgcgggggcgggccctcctgtccaggaggagtgcggcggcctggaggat 2415
          ||| || ||||| |||| ||| ||||||||||||| |||||||||||| ||||||
Db   1083 gagaccgcaatgcggagtcggaccaccagctgcggcgtggagcgcagcctggaggag 1142

Qy   2416 ggggagcctcagcagaagcggggcccacacagagaggagcccaacaggccaagaaggt 2468
          |||  || ||||||| ||||||||||| |||| |||| |||| |||| ||| ||||
Db   1143 gcacggcaggagcggagacgggcggcgggctgaggtgggccgggcagcgcagct 1195
```

```
RESULT  11
ID   AAC38171 standard; DNA; 464 BP.
XX
AC   AAC38171;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 20026.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
```

```
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
```

```
PR  06-MAY-1999;  99US-0132486.
PR  06-MAY-1999;  99US-0132487.
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  14-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  15-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139952.
PR  17-JUN-1999;  99US-0139453.
PR  18-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139954.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142005.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144086.
PR  16-JUL-1999;  99US-0144085.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.

PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  26-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150506.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  30-AUG-1999;  99US-0151080.
PR  31-AUG-1999;  99US-0151303.
PR  01-SEP-1999;  99US-0151438.
PR  07-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0155996.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  08-OCT-1999;  99US-0158369.
PR  12-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
```

RESULT 12
ID AAV55831 standard; DNA; 799 BP.
XX
AC AAV55831;
XX
DT 18-NOV-1998 (first entry)
XX
DE Nucleotide sequence of the stabilising sequence-encoding insert.
XX
KW Fusion protein; stabilising polypeptide; proteolytic degradation;
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;
KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
KW nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
XX
OS Epstein-barr virus.
XX
PN WO982577-A1.
XX
PD 28-MAY-1998.
XX
PF 17-NOV-1997; 97WO-IB01508.
XX
PR 25-JUN-1997; 97US-0048945.
PR 15-NOV-1996; 96US-0030986.
XX
PA (MASU/) MASUCCI M G.
XX
PI Masucci MG;
XX
DR WPI; 1998-312463/27.

PR 21-OCT-1999;
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   2.0%;   Score 50.6;   DB 21;   Length 464;
Best Local Similarity   52.5%;   Pred. No. 0.0023;
Matches 135;   Conservative 0;   Mismatches 119;   Indels 3;   Gaps 1;

```
Qy 1420 gcagagaaaagaagtcagtaccaccagaaatcatctccttggaacaggtctgccatccg 1479
          | |||| | |||| | |        |  ||  ||    |   | ||     |
Db   89 gaaaagattatacgagatgacatgggagattgttattctaggacactgttcgttctcagcct 148

Qy 1480 atgaaagattcgaaatgtcagtgccacacttgtcaaacataagcccgacacgtctctgcta 1539
          ||||| |||  || |   |     | ||      || | || |   |
Db  149 tcaaaataccgaaacgttagtgcaattttattgacttgttctctaggagtagtcttctc 208

Qy 1540 ctggactgtgttggtgaaggggccacatttgggcagctgtgccgtcattacg---agaccaggtg 1596
         || | |||| | |||||| || |||   ||||  ||| |          |||   |  |
Db  209 ctagatggtggaagaaacctgggccagcttaaaagaagatatggttagacggtgca 268

Qy 1597 gacagggtcctgggcaccctgctgctcgtgttcgttgtgtcccacccctgcacgcagagatcacac 1656
         ||  ||  |  ||||  |  ||| || | |  | | |||   |  ||   |||  ||  |  || |
Db  269 gatgaggcggttagaaagttggagatgtatatgggtcctcacatccacgctgatcacca 328

Qy 1657 acgggcttgccaagtat 1673
         || || |||| |  | |
Db  329 acaggtcttgcgaggat 345
```

PT New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
XX
Disclosure; Fig 4B; 120pp; English.

CC This is a nucleotide sequence of the stabilising sequence-encoding
CC insert. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
XX
SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;

Query Match   1.9%;   Score 48.2;   DB 19;   Length 799;
Best Local Similarity   58.9%;   Pred. No. 0.012;
Matches 83;   Conservative 0;   Mismatches 58;   Indels 0;   Gaps 0;

```
Qy 2327 agggaggatggaggagcgcaggggagagcgcggggagcgcaggtgcgggggccctcctgt 2386
         ||||||  |  ||||  || ||||||  ||  ||||  ||||| || |||||||  |  ||
Db  202 aggagcaggaggaggggcaggaggggcaggaggggcaggagg 261

Qy 2387 ccaggggagctggcagcgcggcctggaggatggggaggcctcagcagcagaaagcggcacacag 2446
         |  ||| || || ||| ||||  |||||| ||||||||  | |||| | |   | || ||  ||
Db  262 ggcaggagcaggaggaggggcaggaggggcaggaggggcaggaggggcaggagg 321

Qy 2447 aggaggccacaggccaagaagg 2467
         || ||   ||||| |||| |||
Db  322 aggggcaggaggggcaggagg 342
```

RESULT 13
ID AAX90924 standard; DNA; 1925 BP.
XX
AC AAX90924;
XX
DT 17-JAN-2000 (first entry)
XX
DE Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.
XX
KW Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVEBNA; EBNA 1;
KW episome; transfection; origin of replication; EBV oriP; receptor;
KW eucaryotic host cell; recombinant cell line; ion channel; gene therapy;
KW multiple gene expression; transporter protein; transcription factor;
KW adhesion molecule; antisense therapy; gene amplification;
KW cell immortalisation; ds.
XX
OS Epstein-barr virus.
XX
FH Key   Location/Qualifiers
FH
FT CDS   1..1925
FT   /*tag= a
FT   /product= "EBNA 1"
FT   /transl_except= (pos:799..800, aa:Gly)
FT   /note= "the sequence is described throughout the
FT   specification as being 1926 nucleotides long, but a
FT   sequence of only 1925 bp has been given in figure 2"

```
RESULT 14
ID   AAA50254 standard; DNA; 1926 BP.
XX
AC   AAA50254;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.
XX
KW   EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;
KW   gene therapy; ds.
XX
OS   Epstein-barr virus.
XX
PN   WO200047778-A1.
XX
PN   WO9947647-A1.
XX
PD   23-SEP-1999.
XX
PF   12-FEB-1999;   99WO-US03307.
XX
PR   18-MAR-1998;   98US-0040961.
PR   06-AUG-1998;   98US-0130114.
XX
PA   (PHAR-) PHARMACOPEIA INC.
XX
PI   Damaj BB, Horlick RA, Robbins AK;
XX
DR   WPI; 1999-610610/52.
DR   P-PSDB; AAY28843.
XX
PT   New method for expressing genes from recombinant eukaryotic cells,
PT   useful for gene therapy -
XX
PS   Claim 24; Fig 2; 86pp; English.
XX
CC   The present sequence is a DNA encoding Epstein Barr Virus Nuclear
CC   Antigen 1 (EBNA 1), which is obtained from commercially available
CC   plasmid pCMVEBNA. EBNA 1 protein is used to stably maintain episomes
CC   containing EBV origin of replication (oriP) and a gene encoding
CC   protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
CC   protein are transfected with these episomes to produce recombinant
CC   cell lines expressing multiple genes of interest. This provides a
CC   rapid and reliable method of stably expressing multiple genes in
CC   transfected cells. The episomes are useful in the transfection of genes
CC   encoding receptors, transporter proteins, ion channels, adhesion
CC   molecules and transcription factors. The episomes carrying desired genes
CC   can also be used to transfect cells in gene therapy, antisense therapy,
CC   for gene amplification, cell immortalisation, etc.
XX
SQ   Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;

Query Match             1.9%;  Score 48.2;  DB 20;  Length 1925;
Best Local Similarity   58.9%;  Pred. No. 0.017;
Matches  83;  Conservative  0;  Mismatches  58;  Indels  0;  Gaps  0;

Qy  2327 aggagatggaggaggagcgcaggaggaaagccgggaagctgcgggcgcgccctcctgt 2386
          |||| ||||||||| |  |||| || |||  ||||  || |||  ||||  | |
Db   444 aggagcaggaggaggggcaggaggagcaggaggggcaggaggaggggcaggagg 503

Qy  2387 ccagggagctggccaggggcctgggaggagcctcagcaggagaggagcgccacacag 2446
          ||||||  |||  |||  ||||| || ||  ||  |||||| ||||| || |
Db   504 ggcaggagcaggaggaggggcaggaggaggggcaggaggaggggcaggagg 563

Qy  2447 aggagccacacaggccaagaagg 2467
          |||||  ||  ||| ||||||
Db   564 aggggcaggaggggcaggagg 584
```

```
PD   17-AUG-2000.
XX
PF   11-FEB-2000;   2000WO-US03547.
XX
PR   11-FEB-1999;   99US-0249585.
XX
PA   (PHAR-) PHARMACOPEIA INC.
XX
PI   Horlick RA, Chelsky D;
XX
DR   WPI; 2000-515062/46.
DR   P-PSDB; AAY95856.
XX
PT   Stably transfecting eukaryotic cells with at least one episome for the
PT   production of a desired protein in vitro and for gene therapy -
XX
PS   Disclosure; Fig 2; 53pp; English.
XX
CC   The present sequence is that of DNA encoding the Epstein-Barr virus
CC   (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is
CC   utilised in a novel method for obtaining a eukaryotic cell that is
CC   stably transfected with at least one episome. This method involves
CC   transfecting a eukaryotic cell with: (1) a first episome comprising
CC   an EBV origin of replication (oriP, see AAA50253), a gene encoding a
CC   first protein whose expression results in cell death and a
CC   selectable marker for eukaryotic cells; and (2) a second episome
CC   comprising an EBV oriP and a gene encoding a second protein, where
CC   expression of the second protein prohibits the occurrence of cell
CC   death resulting from expression of the first protein to produce
CC   doubly transfected cells which also express an antigen that
CC   promotes retention of the episomes by the cells. The doubly
CC   transfected cells are maintained under conditions in which the
CC   first and second proteins and the selectable marker are expressed,
CC   and the selective pressure specified by the marker is maintained.
CC   Under these conditions, only cells containing both episomes live.
CC   Preferably, EBNA1 is expressed from 1 of the episomes, and the
CC   protein of interest from the other episome.  Either or both
CC   epitopes may further comprise a nucleic acid sequence encoding a
CC   protein desired to be expressed in the cell (e.g. a therapeutic
CC   protein), a nucleic acid encoding an RNA that is not intended to
CC   be translated (e.g. a therapeutic RNA), or a DNA sequence used as
CC   a tag for the cells.  The method is applicable to cell culture or
CC   intact organisms, for gene therapy.  It allows the rapid
CC   establishment of eukaryotic cells that stably and reliably express
CC   a gene of interest, using a novel method of selection, and
CC   maintenance of that selection without the need for exogenous
CC   selection factors, such as antibiotics.
XX
SQ   Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match             1.9%;  Score 48.2;  DB 21;  Length 1926;
Best Local Similarity   58.9%;  Pred. No. 0.017;
Matches  83;  Conservative  0;  Mismatches  58;  Indels  0;  Gaps  0;

Qy  2327 aggagatggaggaggagcgcaggaggaaagccgggaagctgcgggcgcgccctcctgt 2386
          |||| ||||||||| |  |||| || |||  ||||  || |||  ||||  | |
Db   444 aggagcaggaggaggggcaggaggagcaggaggggcaggaggaggggcaggagg 503

Qy  2387 ccagggagctggccaggggcctgggaggagcctcagcaggagaggagcgccacacag 2446
          ||||||  |||  |||  ||||| || ||  ||  |||||| ||||| || |
Db   504 ggcaggagcaggaggaggggcaggaggaggggcaggaggaggggcaggagg 563

Qy  2447 aggagccacacaggccaagaagg 2467
          |||||  ||  ||| ||||||
Db   564 aggggcaggaggggcaggagg 584

RESULT 15
ID   AAA75454 standard; DNA; 2580 BP.
XX
AC   AAA75454;
```

```
XX
DT   15-JAN-2001  (first entry)
XX
DE   Nucleotide sequence of the Epstein Barr nuclear antigen.
XX
KW   Origin of replication; protein-protein interaction; replication;
KW   two-hybrid system; nuclear antigen; ss.
XX
OS   Epstein-barr virus.
XX
XX
PN   US6114111-A.
XX
PD   05-SEP-2000.
XX
PF   30-MAR-1998;      98US-0050863.
XX
PR   30-MAR-1998;      98US-0050863.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Luo Y,  Payan D,  Huang B;
XX
DR   WPI; 2000-593546/56.
XX
XX
PS   Disclosure; Column 17-20; 18pp; English.
XX
PT   Composition for detecting protein-protein interactions in a mammalian
PT   two hybrid system comprises bait and test vector which consist of
PT   selection gene, vector viral origin of replication and fusion gene  -
XX
CC   The present sequence represents the Epstein Barr nuclear antigen.
CC   It is used to produce bait vectors of the invention. The specification
CC   describes a compositions and methods for a genetic system of detecting
CC   protein-protein interactions in a mammalian host cell. The system
CC   comprises bait and test, both containing selection genes, and viral
CC   origin of replications which require bound viral replication proteins
CC   to effect replication. The compositions is useful for detecting an
CC   interaction between a bait protein and a test protein. It is useful in
CC   a mammalian two-hybrid system for detecting protein-protein interactions
CC   in a mammalian host cell.
XX
SQ   Sequence 2580 BP; 632 A; 512 C; 1054 G; 382 T; 0 other;

Query Match             1.9%;   Score 48.2;  DB 21;  Length 2580;
Best Local Similarity   58.9%;  Pred. No. 0.02;
Matches   83;  Conservative   0;  Mismatches  58;  Indels   0;  Gaps   0;

Qy   2327 aggagattggaggaagcggaggaagcgggagctgcggcagttgcggcgccctcctgt 2386
            ||||| |||||||||| |||||||| || |||||| | |||| || |||| ||
Db    827 aggagcaggaggaggggcaggagcaggaggaggggcaggaggaggggcaggagg 886

Qy   2387 ccagggagctggcagcggccctggaggatggggagctcagcaggaagcgggcccacacag 2446
            ||||||||  |||| |||||||||| ||| || ||| ||||| ||||| ||
Db    887 ggcaggagcaggaggaggggcaggaggaggggcaggaggaggggcaggaggg 946

Qy   2447 aggagccacaggccaagaagg 2467
            ||| ||  |||  ||| |||
Db    947 aggggcaggaggggcaggagg 967
```

Search completed: October 28, 2001, 19:34:56
Job time: 5445 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 17:20:06 ; Search time 66.47 Seconds
                                             (without alignments)
                                             7066.058 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgctttgctcgct..........agaaggtcagagccagtga 2481

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	DB ID	Description
1	58.4	2.4	2517	US-09-315-794-51	Sequence 51, Appl
2	58.4	2.4	2517	US-09-389-341-51	Sequence 51, Appl
3	48.2	1.9	2580	US-08-533-863-2	Sequence 2, Appl
4	48.2	1.9	5452	US-09-130-114-1	Sequence 1, Appl
c 5	48.2	1.9	9600	US-08-910-647-1	Sequence 1, Appl
6	48.2	1.9	10596	US-07-884-811-15	Sequence 15, Appl
7	48.2	1.9	10596	US-07-885-971-15	Sequence 15, Appl
8	48.2	1.9	10596	US-08-087-783A-15	Sequence 15, Appl
9	48.2	1.9	10596	US-08-194-088B-15	Sequence 15, Appl
10	48.2	1.9	10596	US-08-194-087-15	Sequence 15, Appl
c 11	48.2	1.9	10596	PCT-US93-04648-15	Sequence 15, Appl
c 12	43.6	1.8	217	US-09-470-892-3	Sequence 3, Appl
c 13	43.6	1.8	1931	US-09-130-114-2	Sequence 2, Appl
14	42.8	1.7	16442	US-08-781-891-208	Sequence 208, App
15	41	1.7	2301	US-08-306-691B-23	Sequence 23, Appl
16	41	1.7	2301	PCT-US93-06251-78	Sequence 78, Appl
17	40.4	1.6	15378	US-08-785-420-1	Sequence 1, Appl
18	40.2	1.6	1412	US-08-097-831-1	Sequence 1, Appl
19	39.8	1.6	6306	US-08-466-390-3	Sequence 3, Appl
20	39.8	1.6	6306	US-08-470-950-3	Sequence 3, Appl
21	39.8	1.6	6306	US-08-467-781-3	Sequence 3, Appl
22	39.8	1.6	6306	US-08-195-487-3	Sequence 3, Appl
23	39.8	1.6	6306	US-08-483-924-3	Sequence 3, Appl
24	39.8	1.6	6306	PCT-US93-06160-3	Sequence 3, Appl
25	39.6	1.6	2793	US-08-209-747-1	Sequence 1, Appl
26	39.6	1.6	2793	US-08-458-298-1	Sequence 1, Appl
27	39.6	1.6	9551	US-08-056-200-93	Sequence 93, Appl
28	39.6	1.6	9551	US-08-800-644-93	Sequence 93, Appl
c 29	39.6	1.6	51259	US-08-781-891-209	Sequence 209, Appl
30	39	1.6	697	US-17843-10	Patent No. 5171843
31	38.6	1.6	1960	US-08-533-306A-1	Sequence 1, Appli
32	38.6	1.6	1960	US-08-742-923A-1	Sequence 1, Appli
33	38.6	1.6	2680	US-08-533-306A-5	Sequence 5, Appli
34	38.6	1.6	2680	US-08-742-923A-5	Sequence 5, Appli
35	38.6	1.6	2887	US-08-533-306A-3	Sequence 3, Appli
36	38.6	1.6	2887	US-08-742-923A-3	Sequence 3, Appli
37	37.6	1.6	1162	US-08-726-306A-52	Sequence 52, Appli
38	37.4	1.5	1137	US-17843-8	Patent No. 5171843
c 39	37.4	1.5	6085	US-09-029-603-4	Sequence 4, Appli
c 40	36.8	1.5	3480	US-09-226-012-1	Sequence 1, Appli
41	36.8	1.5	3950	US-09-226-012-3	Sequence 3, Appli
42	36.6	1.5	2943	US-08-042-747A-7	Sequence 7, Appli
c 43	36.6	1.5	7218	US-08-232-463-14	Sequence 14, Appl
c 44	36.6	1.5	28958	US-08-258-261B-6	Sequence 6, Appli
c 45	36.6	1.5	28958	US-08-456-837-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-315-794-51
; Sequence 51, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-053
; CURRENT APPLICATION NUMBER: US/09/315,794
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 51
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-315-794-51

Query Match   2.4%; Score 58.4; DB 4; Length 2517;
Best Local Similarity   51.2%; Pred. No. 4.3e-06;
Matches 168; Conservative   0; Mismatches 151; Indels   9; Gaps   1;

Qy   2021 tccggatggggaaagatgccacctcctgatacatgaagccaccctggaagatggttgg 2080
            || || ||| || ||||| | | ||| |||| | |||||| || || | |
Db   2126 tcgaaataggctataattcagatctattaattcacgaagctacactagaaaatcaagctac 2185

Qy   2081 aagaggaggaagcagtgggaaaaagacaacacacagcacaagtcccaagcacaagctgtggggatgc 2140
            ||||| | || ||||||| | |||||||||| | |||||| |||||| | ||| |||| |
Db   2186 tggaggatgccgtgaagagaaaaacactgcactattaatgaagcaatcggtgttcgaaca 2245

Qy   2141 ggatgaacgcggagttcattatgctgaccacttcagcacgccgtatgccaagtccccc 2200
            ||||| || | ||| | | |||| || || ||||| |||| | |||||| ||||
Db   2246 aaatgaatgctaggaagttgatcttaacacactttccagagatatcccaaattgccc 2305

Qy   2201 tcttcagccccaacttcagcgagaaagtgggag--------ttgccttgaccacatga 2251
            | |||| |||||| |||||| | |||| |||| ||||| ||||||| |||
Db   2306 aattagacaacaatatatcgatgtoatgccagagagaattttgtcttcgacagtatga 2365

Qy   2252 aggtcgtcgcttggagacttccaacaatgccaagctgattcccccactgaaagccctgt 2311
            || | | | ||||| | ||| |||||| ||||| |||||||| || | || |
Db   2366 tcgttgattatgagaaaaattggtgaacagcacgctatttttccactgctgaataaggcat 2425

Qy   2312 ttgctggcgacatcgaggagatcgagga 2339
            |||| || || || ||||| || ||||
Db   2426 tcgttgaagaaagaagaagaagaagaaga 2453

RESULT 2
US-09-389-341-51
; Sequence 51, Application US/09389341
; Patent No. 6220803
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-057
; CURRENT APPLICATION NUMBER: US/09/389,341
; CURRENT FILING DATE: 1999-09-02
; EARLIER APPLICATION NUMBER: 09/315,794
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 2517
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-389-341-51

Query Match   2.4%; Score 58.4; DB 4; Length 2517;
Best Local Similarity   51.2%;
Matches 168; Conservative 0; Pred. No. 4.3e-06; Mismatches 151; Indels 9; Gaps 1;

Qy   2021 tccggatgggggaaagatgccaccctcctgatacatgaagccaccctggaaagtggtttgg 2080
              ||| | || |||||| | ||| | || |||||||| |||| | |||||| ||||
Db   2126 tcgaaataggctataatccagatctattaattcacgaagctacactagaaaatcaagctac 2185

Qy   2081 aagagagaagcagctggaaaagacacacgcacccagcgcatcaagcgtggggatgc 2140
              |||| | || | | || || | | || ||| | || | | | ||| |
Db   2186 tggaggatgccgtgaagaaaaaacactgcactattaatgaagcaatcgggtgttccgaaca 2245

Qy   2141 ggatgaacgcggagttcattattgctgaaccacacttcagccagccgtatgccaaggtcccc 2200
              ||| | | | | |||| ||| | ||| | | ||| | || | | || | |||| |||
Db   2246 aaatgaatgctaggaagttgatcttaacacacactttcccagagatatcccaaattgccc 2305

Qy   2201 tcttcagcccccaacttcagcgacgaaaagtggggg----------ttgcctttgaccacatga 2251
              ||||| | || | || | | | | || | | ||||| ||||| | |||
Db   2306 aattagacaatatatattgtatgtgatggcgagagaaattgtcttcgacagtatga 2365

Qy   2252 aggtctgtcttggagactttccaacaatgcccaagctgattccccccactgaaagccctgt 2311
              |||||| | | ||||| ||||| || ||||||||||| | ||| | |||| |||| |||
Db   2366 tcgttgattatgagaaaaattggtgaacacagcagcgtattttccactgctgaataaggcat 2425

Qy   2312 ttgctgtcgcgacatcgaggagatggagga 2339
              ||| | | || | ||||| || ||| |||
Db   2426 ttgttgaagaaaaggaagaagaagaagaaga 2453

RESULT 3
US-09-050-863-2
; Sequence 2, Application US/0905086 3
; Patent No. 6114111
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; APPLICANT: Hiang, Betty
; APPLICANT: Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; TITLE OF INVENTION: System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94111-4187
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   CURRENT APPLICATION NUMBER: US/09/050,863
;   APPLICATION NUMBER: US/09/050,863
;   FILING DATE: 30-MAR-1998
;   CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Silva, Robin M.
;   REGISTRATION NUMBER: 38,304
;   REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2580 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA
US-09-050-863-2

Query Match   1.9%; Score 48.2; DB 3; Length 2580;
Best Local Similarity   58.9%;
Matches   83; Conservative 0; Pred. No. 0.0024; Mismatches 58; Indels 0; Gaps 0;

Qy   2327 agggagggatggaggagcgcgcaggagacggggaagcggggcacgtgcgggggcccctcctgt 2386
              ||||| |||||||||| | | |||||| |||||| | | ||| | | ||||| || ||||
Db   827 AGGACCAGGAGGAGGGGCCCAGGAGGGGCAGGAGGAGCAGGAGGGGGCAGGGGGCAGGAGG 886

Qy   2387 ccagggagctggcaggcgccgctgagaatggggggagatggggaagcggggcccacacag 2446
              || || ||| ||||| | | || | |||||| | ||||| | ||||| || | ||
Db   887 GGCAGGAGCAGGAGGAGGGGGCAGGAGGAGCAGGAGGGGGCAGGAGGGGCAGGAGGAGG 946

Qy   2447 aggagccacaggccaagaagg 2467
              |||| | || ||| |||| ||
Db   947 AGGGGCAGGAGGGGGCAGGAGG 967

RESULT 4
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 5452
;   TYPE: DNA
;   ORGANISM: VEBNA
US-09-130-114-1

Query Match   1.9%; Score 48.2; DB 2; Length 5452;
Best Local Similarity   58.9%;
Matches   83; Conservative 0; Pred. No. 0.0036; Mismatches 58; Indels 0; Gaps 0;

Qy   2327 agggagggatggaggagcgcgcaggagacggggaagcggggcacgtgcgggggcccctcctgt 2386
              ||||| |||||||||| | | |||||| |||||| | | ||| | | ||||| || ||||
Db   1978 AGGACCAGGAGGAGGGGCCCAGGAGGGGCAGGAGGAGCAGGAGGGGGCAGGGGGCAGGAGG 1919

Qy   2387 ccagggagctggcaggcgccgctgagaatggggggagatggggaagcggggcccacacag 2446
              || || ||| ||||| | | || | |||||| | ||||| | ||||| || | ||
Db   1918 GGCAGGAGCAGGAGGAGGGGGCAGGAGGAGCAGGAGGGGGCAGGAGGGGCAGGAGGAGG 1859

```
Qy 2447 aggagcccaagccaagaagg 2467
        ||| ||| ||  || ||||
Db 1858 AGGGGCAGGAGGGGCAGGAGG 1838
```

RESULT 5
US-08-910-647-1
Sequence 1, Application US/08910647
Patent No. 6251433
GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Chiron Corporation
;   STREET: 4560 Horton Street
;   CITY: Emeryville
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 94608-2916
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
;   NAME: Fujita, Sharon M.
;   REGISTRATION NUMBER: 38,459
;   REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510) 923-2706
;   TELEFAX: (510) 655-3542
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 9600 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match   1.9%; Score 48.2; DB 4; Length 9600;
Best Local Similarity 58.9%; Pred. No. 0.0048;
Matches   83; Conservative   0; Mismatches 58; Indels   0; Gaps   0;

Qy 2327 aggagatggagagcgcaggagaagctgcggcaggtcgggcggccctcctgt 2386
        |||||| ||||||| ||||| | ||||| ||||||| ||| || ||||
Db 873 AGGAGCAGGAGAGAGCGGGGGGCCAGGAGGAGGGGGCAGGGGCCAGGAGG 932

Qy 2387 ccagggagctgcggcaggcgcctcagcagaagcgggccccacacag 2446
        ||||| || | || |||| |||| | ||| |||| | ||||||
Db 933 GGCAGGAGCAGGAGAGAGCGGGAGCACAGGAGGAGGGGGCCAGGAGG 992

Qy 2447 aggagcccaagccaagaagg 2467
        ||| || || || ||||
Db 993 AGGGGCAGGAGGGGCAGGAGG 1013
```

RESULT 6
US-07-884-811-15
Sequence 15, Application US/07884811
Patent No. 5316921
GENERAL INFORMATION:
;  APPLICANT: Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;  TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
;  NUMBER OF SEQUENCES: 21

---

;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 460 Point San Bruno Blvd
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: patin (Genentech)
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/884,811
;   FILING DATE: 19920518
;   CLASSIFICATION: 530
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Dreger, Ginger R.
;   REGISTRATION NUMBER: 33,055
;   REFERENCE/DOCKET NUMBER: 755.1
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415/225-3216
;   TELEFAX: 415/952-9881
;   TELEX: 910/371-7168
;  INFORMATION FOR SEQ ID NO: 15:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 10596 bases
;    TYPE: NUCLEIC ACID
;    STRANDEDNESS: single
;    TOPOLOGY: linear
US-07-884-811-15

Query Match             1.9%;  Score  48.2;  DB 1;  Length 10596;
Best Local Similarity  58.9%;  Pred. No. 0.0051;
Matches   83;  Conservative    0;  Mismatches  58;  Indels    0;  Gaps    0;

Qy 2327 aggagatggagagcgcaggagaagctgcggcaggtcgggcggccctcctgt 2386
        |||||| ||||||| ||||| |  |||||  ||||||| ||| || ||||
Db 2407 AGGAGCAGGAGAGAGCGGGGGGCCAGGAGGAGGGGGCAGGGGCCAGGAGG 2466

Qy 2387 ccagggagctgcggcaggcgcctcagcagaagcgggccccacacag 2446
        ||||| ||  | || |||| |||| | |||  |||| |  ||||||
Db 2467 GGCAGGAGCAGGAGAGAGCGGGAGCACAGGAGGAGGGGGCCAGGAGG 2526

Qy 2447 aggagcccaagccaagaagg 2467
        ||| || ||  || ||||
Db 2527 AGGGGCAGGAGGGGCAGGAGG 2547
```

RESULT 7
US-07-885-971-15
Sequence 15, Application US/07885971
Patent No. 5328837
GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 460 Point San Bruno Blvd
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
COMPUTER READABLE FORM:
;   MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/885,971
  FILING DATE: 19920518
  CLASSIFICATION: 530
PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:
ATTORNEY/AGENT INFORMATION:
  NAME: Dreger, Ginger R.
  REGISTRATION NUMBER: 33,055
  REFERENCE/DOCKET NUMBER: 779
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 415/225-3216
  TELEFAX: 415/952-9881
  TELEX: 910/371-7168
US-07-885-971-15

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
  LENGTH: 10596 bases
  TYPE: NUCLEIC ACID
  STRANDEDNESS: single
  TOPOLOGY: linear

Query Match   1.9%;   Score 48.2;   DB 1;   Length 10596;
Best Local Similarity   58.9%;   Pred. No. 0.0051;
Matches 83;   Conservative   0;   Mismatches 58;   Indels   0;   Gaps   0;

Qy 2327 aggagatggaggagcgcaggggagaagcgggagctgcgggtgcgggcggccctcctgt 2386
         ||||| ||| |||||| ||||||| ||||||| ||| | ||| | || |||| |||
Db 2407 AGGAGCAGGAGGAGGGGGCCAGGAGGCGGCAGGAGGAGGGGGCAGGAGGGGCAGGGAGG 2466

Qy 2387 ccagggagctggcaggcgcgcctggaggatggggaggacctcagcagagaagcgggcccacacag 2446
         ||||| |||||| || | | || ||| ||||| | | || | || ||| ||| ||
Db 2467 GGCAGGAGCAGGAGGAGGGGGCAGGAGGAGGGCAGGAGGAGGGGGCAGGAGGGGCAGGGAGG 2526

Qy 2447 aggagccacaggccaagaagg 2467
         ||||| ||| || ||||
Db 2527 AGGGGCAGGAGGGGGCAGGAGG 2547

RESULT 8
US-08-087-783A-15
Sequence 15, Application US/08087783A
Patent No. 5547856
GENERAL INFORMATION:
  APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
  TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
  NUMBER OF SEQUENCES: 22
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Genentech, Inc.
    STREET: 460 Point San Bruno Blvd
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: WinPatin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/087,783A
    FILING DATE: 13-Jul-1993
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/884811
    FILING DATE: 18-MAY-92
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/885971
    FILING DATE: 18-MAY-92
  ATTORNEY/AGENT INFORMATION:
    NAME: Marschang, Diane L.

REGISTRATION NUMBER: 35,600
REFERENCE/DOCKET NUMBER: P0755779P1
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 415/225-5416
  TELEFAX: 415/952-9881
  TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
  LENGTH: 10596 base pairs
  TYPE: Nucleic Acid
  STRANDEDNESS: Single
  TOPOLOGY: Linear
US-08-087-783A-15

Query Match   1.9%;   Score 48.2;   DB 1;   Length 10596;
Best Local Similarity   58.9%;   Pred. No. 0.0051;
Matches 83;   Conservative   0;   Mismatches 58;   Indels   0;   Gaps   0;

Qy 2327 aggagatggaggagcgcaggggagaagcgggagctgcgggtgcgggcggccctcctgt 2386
         ||||| ||| |||||| ||||||| ||||||| ||| | ||| | || |||| |||
Db 2407 AGGAGCAGGAGGAGGGGGCCAGGAGGCGGCAGGAGGAGGGGGCAGGAGGGGCAGGGAGG 2466

Qy 2387 ccagggagctggcaggcgcgcctggaggatggggaggacctcagcagagaagcgggcccacacag 2446
         ||||| |||||| || | | || ||| ||||| | | || | || ||| ||| ||
Db 2467 GGCAGGAGCAGGAGGAGGGGGCAGGAGGAGGGCAGGAGGAGGGGGCAGGAGGGGCAGGGAGG 2526

Qy 2447 aggagccacaggccaagaagg 2467
         ||||| ||| || ||||
Db 2527 AGGGGCAGGAGGGGGCAGGAGG 2547

RESULT 9
US-08-194-088B-15
Sequence 15, Application US/08194088B
Patent No. 5580963
GENERAL INFORMATION:
  APPLICANT: Godowski, Paul J., Lokker, Nathalie A. Mark, Melanie R.
  TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
  NUMBER OF SEQUENCES: 21
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Genentech, Inc.
    STREET: 460 Point San Bruno Blvd
    CITY: South San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94080
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: patin (Genentech)
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/194,088B
    FILING DATE: 09-FEB-1994
    CLASSIFICATION: 530
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/884811
    FILING DATE: 18-MAY-1992
  ATTORNEY/AGENT INFORMATION:
    NAME: Gallegos, R. Thomas
    REGISTRATION NUMBER: 32,692
    REFERENCE/DOCKET NUMBER: 755D1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 415/225-2614
    TELEFAX: 415/952-9881
    TELEX: 910/371-7168
  INFORMATION FOR SEQ ID NO: 15:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 10596 bases
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear

US-08-194-088B-15

```
Qy         Query Match        1.9%;    Score 48.2;    DB 1;    Length 10596;
           Best Local Similarity    58.9%;    Pred. No. 0.0051;
Matches    83;    Conservative    0;    Mismatches    58;    Indels    0;    Gaps    0;

Qy   2327  aggagatggaggagcgcaggggagaacgggagctgccggcgcggcccctccctgt 2386
           |||||  ||||||| |  ||||||  |  ||||  |||| | | |||||||||||
Db   2407  AGGAGCAGGAGGAGCGCAGGGGCGCAGGAGAGCGCAGGAGAGGGGCAGGAGGAGG 2466

Qy   2387  ccaggagctgccaggcgcggcccctcagcagagaagcgggcccacacag 2446
           |||||| ||  |  ||||| ||||| |||| | |||| || ||||||||
Db   2467  GGCAGGACAGGAGGAGCGCAGGAGAGCGCAGGAGAGGGGCAGGAGGAGG 2526

Qy   2447  aggagcccaagaaagg 2467
           |||||  |||  ||||
Db   2527  AGGGGCAGGAGGGGCAGGAGG 2547
```

RESULT 10
US-08-194-087-15
Sequence 15, Application US/08194087
Patent No. 5879910
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,087
FILING DATE: 18-MAY-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 779
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-194-087-15

```
           Query Match        1.9%;    Score 48.2;    DB 2;    Length 10596;
           Best Local Similarity    58.9%;    Pred. No. 0.0051;
Matches    83;    Conservative    0;    Mismatches    58;    Indels    0;    Gaps    0;
```

```
Db   2467  GGCAGGAGGAGCGCAGGAGAGCGCAGGAGAGGGGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAGGGGCAGGAGGAGG 2526
           |||||| ||| || |||| |||| ||||| || ||| |||| || ||| |||| || || ||| || ||||| || ||
Qy   2447  aggagccacaggccaagaaagg 2467
Db   2527  AGGGGCAGGAGGGGCAGGAGG 2547
```

RESULT 11
PCT-US93-04648-15
Sequence 15, Application PCT/US9304648
GENERAL INFORMATION:
APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melani
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04648
FILING DATE: 19930517
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/885971
FILING DATE: 18-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/884811
FILING DATE: 18-MAY-92
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755,779P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
PCT-US93-04648-15

```
Qy         Query Match        1.9%;    Score 48.2;    DB 5;    Length 10596;
           Best Local Similarity    58.9%;    Pred. No. 0.0051;
Matches    83;    Conservative    0;    Mismatches    58;    Indels    0;    Gaps    0;

Qy   2327  aggagatggaggagcgcaggggagaacgggagctgccggcgcggcccctccctgt 2386
           |||||  ||||||| |  ||||||  |  ||||  |||| | | |||||||||||
Db   2407  AGGAGCAGGAGGAGCGCAGGGGCGCAGGAGAGCGCAGGAGAGGGGCAGGAGGAGG 2466

Qy   2387  ccaggagctgccaggcgcggcccctcagcagagaagcgggcccacacag 2446
           |||||| ||  |  ||||| ||||| |||| | |||| || ||||||||
Db   2467  GGCAGGAGGAGCGCAGGAGAGCGCAGGAGAGGGGCAGGAGGAGG 2526

Qy   2447  aggagccacaggccaagaaagg 2467
           |||||  |||  ||||
Db   2527  AGGGGCAGGAGGGGCAGGAGG 2547
```

RESULT 12
US-08-470-892-3/c

```
  NUMBER OF SEQ ID NOS: 36
  SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2
  LENGTH: 1931
  TYPE: DNA
  ORGANISM: EBNA
US-09-130-114-2

Query Match        1.8%;  Score 43.6;  DB 2;  Length 1931;
Best Local Similarity   54.3%;  Pred. No. 0.036;
Matches   88;  Conservative    0;  Mismatches  74;  Indels    0;  Gaps    0;

Qy  2317 ggcggacatcgaggagagatggaggagcgcgcaggaggaaggaagcgggagctgaaggacaggtcgggccg 2376
              || |  || |  |  || ||||   | ||||   | | ||  |||| | | || | | |||
Db   613 GACGAGGAGACGGGGAGGACGGGGAGGAGGACGCGGGGAGGACGCGGGGAGGAGGACGAG 554

Qy  2377 gccctccctgtccaggagagagctggcaggcggccctggaggatggggaggacgcctcagaagaagcgg 2436
Db   553 GACGGGGAGGACGGGGGGTGGTAGCCCGGAGGACGGGGAGGACGGGGAGGAGGACGCGGGGAGGACGAG 494

Qy  2437 gcccacacaggaggagcccacaggccaagaaggttcagaggcccag 2478
Db   493 GAGGACGGGGAGGAGGACGAGGACGAGGACGGGGAGGAGGACGGGGAGGAG 452

RESULT 14
US-08-781-208/c
Sequence 208, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
  APPLICANT: Fu, Ying-Hui
  APPLICANT: Yu, Chang-En
  APPLICANT: Oshima, Junko
  APPLICANT: Mulligan, John T.
  APPLICANT: Schellenberg, Gerald D.
  TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
  TITLE OF INVENTION: WERNER'S SYNDROME
  NUMBER OF SEQUENCES: 209
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: SEED and BERRY LLP
    STREET: 6300 Columbia Center, 701 Fifth Avenue
    CITY: Seattle
    STATE: Washington
    COUNTRY: USA
    ZIP: 98104-7092
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/781,891
    FILING DATE: 27-DEC-1996
    CLASSIFICATION: 800
  ATTORNEY/AGENT INFORMATION:
    NAME: No. 6090620tenburg Ph.D., Carol
    REGISTRATION NUMBER: 39,317
    REFERENCE/DOCKET NUMBER: 240052.419
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (206) 622-4900
    TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 208:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 16442 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
US-08-781-891-208

Query Match        1.7%;  Score 42.8;  DB 3;  Length 16442;
Best Local Similarity   56.3%;  Pred. No. 0.18;
```

```
Sequence 3, Application US/08470892
Patent No. 5707839
GENERAL INFORMATION:
  APPLICANT:
  APPLICANT: DENOYA, CLAUDIO D., KIM J.
  APPLICANT: STUTZMAN-ENGWALL, KIM J.
  TITLE OF INVENTION: GENES ENCODING BRANCHED-CHAIN
  TITLE OF INVENTION: ALPHA-KETOACID DEHYDROGENASE COMPLEX FROM STREPTOMYCES
  TITLE OF INVENTION: AVERMITILIS
  NUMBER OF SEQUENCES: 14
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: DR. PETER C. RICHARDSON, PFIZER INC
    STREET: 235 EAST 42ND STREET, 20TH FLOOR
    CITY: NEW YORK
    STATE: NEW YORK
    COUNTRY: U.S.A.
    ZIP: 10017-5755
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/470,892
    FILING DATE: 6-JUN-1995
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: SHEVKA, ROBERT F.
    REGISTRATION NUMBER: 31,304
    REFERENCE/DOCKET NUMBER: PC8529B
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212)573-1189
    TELEFAX: (212)573-1939
    TELEX: N/A
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 217 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
US-08-470-892-3

Query Match        1.8%;  Score 43.6;  DB 1;  Length 217;
Best Local Similarity   55.2%;  Pred. No. 0.012;
Matches   85;  Conservative    0;  Mismatches  69;  Indels    0;  Gaps    0;

Qy   69 gcaggcaccgcccgcccgcgaggcggcgccgcaaggaccccgcttgcggcaccgcgaacgcg 128
          || || ||   ||  |   |||   | ||||| | || ||||| || |||| | |||
Db  183 GCGAGCGGTCGACGGCAnCGAGCACCCGGTCAAGGCCCGGAGAAGGTACTCCTTCCTTCCACAC 124

Qy  129 agagaagcgcggaccgtcgggtgtctccggggcggcccaaacaccgtgtaccctgcaaggtggt 188
Db  123 CGCCCGCGGCGGATACGGGCGGCGTGGTAGCCCCGCCCACCCTCAGCACGGTGCCCTCCAGGTGGT 64

Qy  189 ggcagcgcggtagcccgggactcgggcgccgcgcctc 222
Db   63 AGAAGCAGCGCTTCCGTGATGCGCGCGCGATCTC 30

RESULT 13
US-09-130-114-2/c
Sequence 2, Application US/09130114
Patent No. 5976807
GENERAL INFORMATION:
  APPLICANT: Horlick, Robert A.
  APPLICANT: Damaj, Bassam B.
  APPLICANT: Robbins, Alan K.
  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
  TITLE OF INVENTION: From Multiple Transfected Episomes
  FILE REFERENCE: 0867/1D903US1
  CURRENT APPLICATION NUMBER: US/09/130,114
  CURRENT FILING DATE: 1998-08-06
```

```
Matches     80;  Conservative     0;  Mismatches     62;  Indels     0;  Gaps     0;

Qy   2326 gaggagatggaggagcgcagggagaaagcgggagctgcggcgaggtgcgggcggccctctg 2385
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16408 GAGGAGGAGGAGCGCAGGGAGAAAGCGGGAGCTGCGGCGAGGTGCGGGCGGCCCTCTG 16349

Qy   2386 tccagggagctggccaggcggccggagatggggagcctcagagaagcgggcccacaca 2445
          |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16348 GAGCAGGAGGAGCGACAGGAGGAGGAGCGACAGGAGGAGGAGGAGCAGGAGGAGAGCAG 16289

Qy   2446 gaggagccacaggccaagaagg 2467
          ||||||  |||| |||| |||
Db  16288 GAGGAGGAGCAGGAGCAGGAGG 16267
```

```
RESULT 15
US-08-306-691B-23
Sequence 23, Application US/08306691B
Patent No. 5734039
GENERAL INFORMATION:
APPLICANT: Calabretta, Bruno
APPLICANT: Skorski, Tomasz
TITLE OF INVENTION: ANTISENSE
TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
  STREET: Two Penn Center, Suite 1800
  CITY: Philadelphia
  STATE: Pennsylvania
  COUNTRY: U.S.A.
  ZIP: 19102
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
  COMPUTER: IBM PS/2
  OPERATING SYSTEM: MS-DOS
  SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/306,691B
  FILING DATE: September 15, 1994
  CLASSIFICATION: 514
PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:
ATTORNEY/AGENT INFORMATION:
  NAME: Monaco, Daniel A.
  REGISTRATION NUMBER: 30,480
  REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (215) 568-8383
  TELEFAX: (215) 568-5549
  TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
  LENGTH: 2301 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
US-08-306-691B-23
```

```
Qy   2446 gaggagccacaggccaagaaggtcagagccag 2478
          |||| |||||| |||| ||| || ||| ||| ||
Db   140 GAGCAGGAGGAGCGACAGGAGCCAGGAGGAGGAGAGGAG 172
```

```
Search completed: October 28, 2001, 19:32:53
Job time: 7967 sec
```

```
Query Match          1.7%;  Score 41;  DB 1;  Length 2301;
Best Local Similarity  54.2%;  Pred. No. 0.2;
Matches    83;  Conservative    0;  Mismatches    70;  Indels    0;  Gaps    0;

Qy   2326 gagggagatggaggagcgcagggaaagcggagctgcgcgcaggtgcggggcggcctcctg 2385
          |||||  |  |||| |||||  |||||||  |||| | ||  ||| ||||||| |||
Db     20 GGGAGGAGGAGCACAGGAGGAGCGACAGGAGGAGGAGGAGCAGGAGGAGGAGCAGGAGG 79

Qy   2386 tccaggagctgccaggcggcctggaggatgggggagcctcagagaagcgggcccacaca 2445
          |||  ||  | ||||  | |  ||| |  |||  |  |  |  ||  ||||  |||||
Db     80 GAGCAGGAAGCACAGGAGGAGCGACAGGAGAAGGAGGAGGAGCAGGAGAAGCAGGAGGAG 139
```

RESULT 4

```
Db   361 GGCTGAGAATCTCCAGTCTCCTGCAGTTGGAAAGATTGATCAGTTCGCTGTTGCCAACATGTGA
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1899 tttggaagatttcaagacctgtcttggtgcgcactgcaagtgcgtttggctggtct        1958
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 -TTGGAAGAGTTTCAGACCTGTCTTGGTGCGCGGCACTGCCAAGCAATGCGTTTGGCGTGCGCCT 479
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1959 ggtgcacaccctgtggctgaagtggtctattccgggggaccaccaatgccctgcggaggtct  2018
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 GGTGCACACCTCTGGCTGTGGAAAGTGGTCTATTCCGGGACACCATGCCCTCCGAGGCCTCT    539
         |||||||||||||||||||||||||||||||||||||||||||||
Qy  2019 ggtccggatgggaagaatgccaccctcctgatacatgaagcacacccctgggaagatggttt   2078
         |||||||||||||||||||||||||||||||||||||||||||||
Db   540 GGTCCGGATGGGAAAGATGCCACCCTCCTGATACATGAAGCACACCCTGGGAAGATGGTTT     599
         |||||||||||||||||||||||||||||||||||||||||||
Qy  2079 ggaagaggaagcagtggaaaagaacacaagcgtcccaagccatcacggtgggggat         2138
         |||||||||||||||||||||||||||||||||||||||||||||
Db   600 GGAAGAGGAAGCAGTGGAAAAAGAACAACACAGCGTCCCAAGCCATCACGGTGGGGGAT      659
         |||||||||||||||||||||||||||||||||||||||||||||
Qy  2139 gcggatgaaccgcggagttcattatgctgaaccacactccagccagcgcgctatgccaaggtccc 2198
         |||||||||||||||||||||||||||||||||||||||||||||
Db   660 GCGGATGAACCGCGGAGTTCATTATGCTGAACCACACTTCAGC--AGGCTATGCCAAGGT-CC    716
         |||||||||||||||||||||||||||||||||||||| |   |  || |
Qy  2199 cctcttcagcccacttcagcgagagaaagtggagagttgcct   2240
         |||||||||    | | |||||
Db   717 CCTCTTCAGCCCAACTTCAAGCGAGAAAAGTCGGGAGTTGCTTT 758
```

RESULT 4

RESULT 5

```
Query Match              28.3%;    Score 702;  DB 153;  Length 827;
Best Local Similarity    97.8%;    Pred. No. 6e-173;
Matches    796;  Conservative   0;  Mismatches  10;  Indels   8;  Gaps  8;

Qy   686 gccagagaaggaggggtcaggggactctctccctggtcgtagctccatcgtaagcttcact  745
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GCCAGAGAAGGAGGGGTCAGGTACTCTTCCCTGGTCGTAGCTTCATCTGTAAGCTTCACT   60
Qy   746 taaagagaggaaaacttcttgttgctcaaagcaaagcaaagaggggcctcccagtgggacag  805
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TAAAGAGAGGAAAAACTTCTTGGTGCTCAAAGCAAAGCAAAGAGGAGATGGGGCCTCCCAGTGGGACAG  120
Qy   806 ctgccatcgctcccatcattgctgctgtcaaggacgggaaaagcatcactcatgaaggaa   865
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCATCACTCATGAAGGAA  180
Qy   866 gagagattttggctgaagagctgtgtgtactcctccagatcctgtgctgctttgtggtgg  925
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGGTGCTGCTTTGTGTGGTGG  240
Qy   926 tagaaatgtcccagatgaaagcttcaaccaccatcctgtgagaatgtccactttcaggt  985
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TAGAATGTCCAGATGAAAAGCTTCAACCACCATCTGTGAGAATGCCACTTTCAGAGGT  300
Qy   986 accaaggaaaggcagaatgccccccgtggcccttggtggttcacatggcccagccagctgc 1045
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACCAAGGAAAGGCAGAATGCCCCCCGTGGCCCTTGGTGGTTCACATGGCCCAGCCAGCTGTGTC  360
Qy  1046 ttgtggacagcaggtaccagtcagtggagaggtttgggcctgaacaccagcaacttgg 1105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGTGGACAGCAGGTACCAGTCAGTGGAGAGGTTTGGGCCTGAACACCAGCAACTTGG  420
Qy  1106 tcctgaatggaagaactgtgcctcagttcacaacacttcgcagccaccaagaattcaaaacccagc 1165
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TCCTGAATGGAACAGTGTGCCTCAGTTCACAACCTTCGCAGCCACCAAGAATTCAAAACCCAGC  480
Qy  1166 tcaacctcatccaccggacacatcttcccctgtccaccagttccgctgtaagaaggagg 1225
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TCAACCTCATCCACCGGACACATCTTCCCCCTGTCCACCAGTTCCGCTGTAAGAAGGAGG  540
Qy  1226 gcccaccctcagtgtgcccat-ggttcaggtggaatggccctcctcaagttccagctccgt 1284
         |||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||
Db   541 GCCCCACCCTCAGTGTGCCCATGGGTTCAGGGTGAAATGGCCTCCTTCAAGTACCAGCTCCGT  600
Qy  1285 ccaggagggagggtggtcggcacatta-ttactgcaatcctgggggaattcaatgt 1343
         |||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db   601 CCCAGGAGGGAGTGGCAGAGGGATGCCATTATTTACTTGCAAATCCTGAGGAATTCAATAGT  660
Qy  1344 tga-ggcgtgtcagcttccc-aacttccagcagagccgtgcaggagagtaca-ggaggagtgc 1400
         ||| ||||||||||||||||  ||||||||||||||||||||||||||||||  ||||||||||
Db   661 TGAGGGCGGCCTGCAGCTTCCCAAACTTCCAGCAGCAGCCGTGCCAGGAGTACAGGGGAGGAGTGC  720
Qy  1401 gcaggacggcccagcccccagccagagaaaaagtcaagtaccc-agaaatcaatcttccctg 1459
         ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db   721 GCAGGAACCGGCCAG-CCCAGCCAGAGAAAGAAGTCAGTACCCAAGAAAAATCATCTTTCTTG  779
Qy  1460 g-aacagggtctgccatcccgataagttcgaa 1492
         | |||||||||||||||||||||||||||| ||
Db   780 GAAACAGGGTTCTGGCCATCCGGATGAAGATCGAAA  813
```

RESULT 5

```
RESULT 3

LOCUS       BE795820        761 bp    mRNA        EST      20-SEP-2000
DEFINITION  601509856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
            mRNA sequence.
ACCESSION   BE795820
VERSION     BE795820.1  GI:10217018
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 761)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM802 row: P column: 14
            High quality sequence stop: 761.
FEATURES             Location/Qualifiers
     source          1..761
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3945085"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
BASE COUNT    172 a    206 c    219 g    164 t
ORIGIN

Query Match             28.6%;  Score 709.6;  DB 140;  Length 761;
Best Local Similarity   98.3%;  Pred. No. 6e-175;
Matches  749;  Conservative  0;  Mismatches  9;  Indels  4;  Gaps  3;

Qy   1479 gatgaagagattcagatgcccacacttgtcaacataagccccgacaacgtctctgct 1538
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      1 GATGAAGAGATTCAGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACAACGTCTCTGCT 60

Qy   1539 actggactgtggtgagggcacattgggcagctgtgccgtcgtccgtcattacggagaccaggtgga 1598
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     61 ACTGGACTGTGGTGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGGACCAGGTGGA 120

Qy   1599 cagggtcctgggcaaccctgctgctgttgttgtccccacctgcacgcagatcacccac 1658
          ||||||||||||||||  ||||||| ||||||||||||||||||||||||||||||||
Db    121 CAGGGTCCTGGGCAACCTGCTGGCACGTGCTGTTTGTGTCCCCACCTGCACGCAGATCACCAC 180

Qy   1659 gggcttgccaagtatcttgctgcagagagaacgcgcccttggcatcttcttgggaaagccgt 1718
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    181 GGGCTTGCCAAGTATCTTGCTGCAGAGAGAACCGCGCCCTTGGCATCTTCTTGGGAAAGCCGT 240

Qy   1719 tcaccctttgctggtgttgtcccccaaccagctcaaagccctggctccagcagtaccacaca 1778
          |||||||||||||||||||| |||| ||||||||||||| |||||||||||||||||||||
Db    241 TCACCCTTTGCTGGTGTTGCCCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACA 300

Qy   1779 ccagtgccaggaggtcctgcaccacatcagttatgatccctgccaaatgccttcaggaagg 1838
          ||||||||||||||||||||||||| |||||| |||||| |||||||||||||||||||||
Db    301 CCAGTGCCAGGAGGTCCTGCACCACCATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGG 360

Qy   1839 ggctgagatctccagtcctgcagtggaaagattgatcaagttcgctgttgtcgaacatgtga 1898
```

```
BASE COUNT    230 a    286 c    276 g    168 t
ORIGIN      Note: this is a NIH_MGC Library."

Query Match             29.18%;  Score 721.2;  DB 169;  Length 960;
Best Local Similarity   94.7%;  Pred. No. 5.8e-178;
Matches  801;  Conservative  0;  Mismatches  38;  Indels  7;  Gaps  5;

Qy   1362 caacttcccagcaggaggcgtgcaggagtacaggaggaggggtcgccaggagcgccccagc 1421
          ||||||||||||||||||||||||||||||||||||||  |||  |  ||||||||||||
Db      1 CAACTTCCAGCAGCAGCAGCGTGCAGGAGTACAGGAGGACGGTCCGCCAGGCCCAGC 60

Qy   1422 agagaaaagaagtcagtacccagaaatcatctcttccttggaacagggtctgccatcccgat 1481
          |||||||||||||||||||||||||||||||||  |||||||||||||||| ||| ||||||
Db     61 AGAGAAAAGAAGTCAGTACCCAGAAATCATCTTCCTTGGAACAGGGTTCTGCCCATCCGAT 120

Qy   1482 gaagattcgaaatgtcagtgccacacttgtcaacataagccccgacaacgtctctgctact 1541
          |||||||||||||||||||||||||||||||||||| |||||||||||||||| ||||||
Db    121 GAAGATTCGAAATGTCAGTGCCACACTTGTCAACATTAAGCCCCGACAACGTCTTGCTACT 180

Qy   1542 ggactgtgtgtgaggggcacacattgggcagctgtgccgtcgtgtgtcattacggagaccaggtggacag 1601
          ||||||||| |||||| |||  ||||||||||||||||||||||||||||||||||| |||||||||||
Db    181 GGACTGTGGTGTGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGGACCAGGTGGACAG 240

Qy   1602 ggtcctgggcaaccctggctgctgttgtgtccccacctgcacgcagatcaccacacggg 1661
          |||||||||||||||||||||| | |||| | ||||||||||||||||||||||||||||
Db    241 GGTCCTGGGCAACCCTGGCTGTCGTGTTTGTTGTCCCACCTGCACGCAGATCACCACACGGG 300

Qy   1662 cttgccaagtatcttgctgcagagagaaccgcgccccttggcatcttcttgggaaagccgttca 1721
          |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db    301 CTTGCCAAGTATCTTGCTGCAGAGAGAACCGCGCCCTTGGCATCTTCTTGGGAAAGCCGTTCA 360

Qy   1722 ccctttgctgggtgttgtccccaccagcatcaaagccctggctccagcagtaccacaacca 1781
          ||||||||||| ||||| ||||||||||  ||||||| |||||||||||||||||||| |||
Db    361 CCCTTTGCTGGTGTTGCCCCACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCA 420

Qy   1782 gtgtccaggaggtcctgcaccaccatcagtgatgattcctgccaaatgccttcaggaagggggc 1841
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    421 GTGCCAGGAGGTCCTGCACCACCATCAGTGATGATTCCTGCCAAATGCCTTCAGGAAGGGGC 480

Qy   1842 tgagatctccagtcctgcagtggaaagatgatcagttcgctgttgtcgaacatgtgattt 1901
          |||||||||||||||||||||||||||| | |||||||||||| || |||||||| |  ||
Db    481 TGAGATCTCCAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGTTGCGAACATGTGAATCT 540

Qy   1902 ggaagagttcagacctgtctggtgtgcgggcactgcaagcatgcgtttggctgtcgtgctggt 1961
          |||||||| ||||||||||||||| ||||||||||||||||  | ||||||||| | ||||||
Db    541 GGAAGAGTTTCAGACCTGTCTGGTGCGGGCACTGCAAGCATCGATTGGCTGTGCGCTGGT 600

Qy   1962 gcacacctctggctggaaagtgttctattccgggacaccatgtgccctgccctcggct 2021
          |||||||||||||| |||||||| ||||||||||||||||||| |||||||| ||| |||
Db    601 GCACACCTCTGGCTCGGGAAAGTGGTCTCTATTCCGGGACACCATGCCCTGCCAGGCTCTGGT 660

Qy   2022 ccggatggggggaaagatgccaccacctgataacctggaagagatggttggga 2081
          ||||||||||| ||||||||||  |||  ||||  | |||   ||   ||||||  |||
Db    661 CCGGATGGGGGAACGATGCCA--CCTCCTGATTCATGAAGC--ACCTGGAAGATGGTACGGA 717

Qy   2082 agaggaagcagt-ggaaaagcaacaacagcacaagtccaaagccatcaagcg-tggggatg 2139
          ||||||||||   ||  | |||| ||| |  ||| |||||| ||||  ||| | |||||||
Db    718 AGAGGAAGCACTGGCACAGCCCGCCAACGTCCCAAGGCATCAGCCGGGGGATC 777

Qy   2140 cggatgaacgc--ggagttcattatgctgaaccacttcagcagccgctatgccaaggtcc 2197
          |||||||||| | ||   ||| |  |  ||  | | || |||||  | |   |||| |||
Db    778 CGGATGAACGCCGAGCGTCACAATAGGCGGGAACCAGTCCAGCACGGCTAGGCAAAGGCCC 837

Qy   2198 ccctct 2203
          | | |
Db    838 CAATAT 843
```

```
ORIGIN
BASE COUNT      373 a     412 c     414 g     311 t

        polyA_site      1510
                        /note="putative"
        polyA_signal    1488..1493
                        /note="putative"

                        ALLTQQADSPEDREPQQKRAHTDEPHSPQSKKESVANTLGARV"
                        FSPDFNEKVGIAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVRERKRELRLVRA
                        GKDATLIHEATLEDGLEEEAVEKTHSTTSQAINVGMRWNAEFFIMLNHFSORYAKIPL
                        TTLERLISLLLETCDLEEFQTVCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQM
                        QREHALASLGKPFQPLLVVAPTQLRAWLQQYVHNHCCEILHIVSMIPAKCLQKGAEVSN
                        LVNLSPDKSVLLDCGBGTFCGOLCRHVGOQIDRVLGLSLTAVPVSHLHADHHRGLLNIL
                        EFIAEALELPSFQESVEYEYKNVQQENPAPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSST
        /translation="SSFYSKEGSTILSVPFIVRGECLLKYQLRPKREWQRDTTLDCNTD
        /db_xref="GI:12835202"
        /protein_id="BAB23185.1"
```

```
Query Match          37.0%;  Score 917.4;  DB 192;  Length 1510;
Best Local Similarity  83.7%;  Pred. No. 2.5e-229;
Matches 1051;  Conservative   0;  Mismatches 201;  Indels   3;  Gaps   1;
```

```
Qy  1202 ccagttccgctgtaagaaggaaggagggcccaacctcagtgtgtgcccatgttcaggtgaat 1261
         |||||| || || || |||||||||||||||||||||||||||||||| |||  ||  ||||
Db     2 CCAGCTTCTAPAGTAAGGAGGAGGAAGGGTCCACCCTCAGGGTGCCAACAGTTCGGGTGAAT   61

Qy  1262 gcctcctccagtaccacgctccgtccgtccccagagggaggaggatggccattattactt 1321
         ||||||||||||| |||||| |||| ||||||| ||||||| | ||||||||||||| ||||
Db    62 GCCTCCTCAAGTATCAGCTCCGCCCCAAGAGAGAGTGGCAGAGAGTGGCCACCACTCGACT  121

Qy  1322 gcaatcctgaggaattcaatagttgaggcgtcgcagttcccaacctccagagaggagtgc 1381
         |||||||||| ||||||||| ||||||||| ||||| ||| |||||||||||| || |||
Db   122 GCAATACTGATGAAATTCAATAGCTGAGGCCTTGAGCTCCCCAGTTCCAGGAGAGTGTGG  181

Qy  1382 aggagtacagaggagtgcgcaggacgacggccagccagcagagaaaagaagtcagtacc 1441
         |||||||||||||||||||||||||||||||||||||||||||||| ||||| || |||
Db   182 AGGAGTATCGGAAGAACGTGCAGGAAAAACCCAGCCAGCAGAAAAGAAGCCAGTATC  241

Qy  1442 cagaaatcatctcctggaacaggggtctgccaatcccgatgaagattcgaaatgtcagtg 1501
         ||  |||||||||  ||||| || |||| |||||||| ||||||||| |||||| ||||||
Db   242 CTGAAATTGCTTCCTGGGTACGGGGTCTGCCCATCCAATGAAGATCCGAAAATGTCAGTT  301

Qy  1502 ccacacttgtcaacataaagccccgacacgtctctgtactgtgttggttgaggacaat 1561
         |||||||||||| |||| |||| || || ||||||||||| |||| |||| || ||||||
Db   302 CCACACTCGTCAACCTTAAGCCCCTGACAAGTCAGTCGTCCTGATTGTGGAGGAAGGCACTT  361

Qy  1562 ttgggcagctgtgccgtcattacggagaaccaggtggacaaggtgtcctgggcaccctg 1621
         |||||||| || |||| ||||||| ||||||||||| || |||| ||||||| |||||| |
Db   362 TTGGGCAGTGTGTGCCCTCATTACGGACAGCAAATAGACCGAGTCTTATGCAGCCTCACG  421

Qy  1622 ctgtgtttgtgtccccaactgcagcagatcaccacacgggggctttgccaagtatcttgtgc 1681
         |||||| |||| ||||||||||||| || |||||||||||||||| ||||| |||||| ||
Db   422 CTGTGTTTGTGTCCCCAACCTPAAGGCCTGCACACACGGGGTTGCTGAATATCTTGCTGC  481

Qy  1682 agagagaaacggcctggaaaacgccgttcaccctttgctctggttgttgtgcc 1741
         ||||||||||| ||||| ||  ||||||||| || |||||||  ||||| ||||
Db   482 AGAGAGACATCCGTTGGCACTCTCGGGAAACCCTTCCAGCCCTTGCTCTTGTGGCTGCCTC  541

Qy  1742 ccaacaagctcaaagcctggttccagcagtaccacacagtgccaggagggaggtcctgcacc 1801
         ||||| ||||||||| ||||| ||||| || |||||||  || ||||| |||||| ||||||
Db   542 CTACCCAGCTCAGGGCCTGCGCTCCACCAGTACACACAACCAGGGAGGTTCTGCACC  601

Qy  1802 acatcagtatgattcctgccaaatgcctcagggaagggctgagatctccagtcctgcag 1861
         ||  | ||  |||||| ||||||||||||||| || ||||||||||||| |||||||||||
Db   602 ACGTCAGTAAGATTCCTGCCAAATGCCTTCAGAAAAGGGCAGGGTCTCAATACTACAT  661

Qy  1862 tggaaagaatggatcagttcgtggaagagtttcagacctgtc 1921
         |||| | || || || ||| ||||||||| | |||||||| ||
Db   662 TGGAAAGGCTGATAAGCTTGCTGTTTGGAAAACATGTGACTTAGAAGAATTTCAGACCTGCC  721

Qy  1922 tggttgcggcactgcaagcatgcgttggtgcaacacctctgctggaaag 1981
         |||||||||||||||||||||| ||||||| ||||||||| |||||||||
Db   722 TGGTACGGCACTGCAAGCACATCTTTTGGCTGTTGCACTTCAATCTGCTGGAAAG  781
```

```
Qy  1982 tggtctattccgggaccaccatgccctgccagcgctctggtccggatcggggaaagaatgcca 2041
         |||||  || ||| |||||| |||| |||||||||||||||||| ||||||| |||||||||||
Db   782 TCGTCTACTGCGGGGGATACCATGCCCTGCCCTGTGAGCGCTCTGGTCCAGATGGGGAAAGAATGCCA  841

Qy  2042 cccctcctgatacatgaagcccaccctggaagatggtttggaagaggaaggagagcagtggaaaaa 2101
         |||||| ||| || ||||||  || |||||| || ||| | |||||||||| ||||| || |||||
Db   842 CCCTCCTGATACATGAAGCCACACTCTGGACGATGCTTGGAAGAGGAAGCAGTAGCAGGAAGAGA  901

Qy  2102 cacaacagcacaagtccaagccaagccatcaagcgtgtgggggatcgggagaaccggagttcatta 2161
         |||||| | || | |||||||||||| |||||||||| |||| || |||||| |||||| |||  |
Db   902 CACACAGCACCACCTCCCAGGCTPATTAATGTGGGGATGGGGATGGGGAGTTCATCA  961

Qy  2162 tgctgaaccacttcaagcacgctatgccaagtgtccccctcttcagcccaaacttcagg 2221
         | |||||||||||||| ||||| ||||| |||||||| | || |||| ||||||||| ||
Db   962 TGCTGAAACCACTTCAGTCACGCGTACGGCCAAGAGCCCCTTTTTCAGCCCTGACTTCAACG 1021

Qy  2222 agaaagtgggagttgccttgaccacatgaagtctgtcttttggagacttccaacaatgc 2281
         || ||| || || ||||||| ||||||||||||| ||| |  || || |||||| ||||||
Db  1022 AGAAAGTTCGGCATCGCCTTTGACCACATGAAGGTCTGCTTTGGAGGACTTCCCGACAGTGC 1081

Qy  2282 ccaagctgattccccactgaaagccccgtttgctgctgcgcagatagagagggaggagc 2341
         ||||||||| ||||| ||||||||| ||||||||| | |||| ||||| ||||||||||||
Db  1082 CCAAGCTGATTCCCCACTGAAAGCCCCGTTTGCTGCTGCACATTGAAGAGATGGTGGAAC 1141

Qy  2342 gcaggggagaaagcgggagctgcggcggggcgggcccctcctcgtccagggggagctggcag 2401
         |||| | ||||| ||||||||| |||||| |||||||||| |||||| |||||||||||||||
Db  1142 GCAGGGGAGAAAGGGGAGCTTACGGCGTGCGGGCGGGCCCTTCCCACAGGGAGCTGGCAG 1198

Qy  2402 gcggcctggaggatgggaggcctcagcaagcggggccccacacagaggagcaca 2456
         || | |||||||||| |||| ||||| |||||||||||||||||||||||||| ||
Db  1199 ACACCCCAGGAGCACAGGAGCAGGAGCGAGGAGAACCCCACAACACAGATGAACCACA 1253
```

```
RESULT  2
BF797306
LOCUS       BF797306      960 bp    mRNA      EST      12-JAN-2001
DEFINITION  602256926F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4340286 5',
            mRNA sequence.
ACCESSION   BF797306
VERSION     BF797306.1  GI:12102360
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM952 row: c  column: 07
            High quality sequence stop: 705.
FEATURES             Location/Qualifiers
     source          1..960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4340286"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
```

```
Db   723 TGGTACGGCACTGCAAGCACATCTTTTTGGCTGTTGCACTTCAATCTGCTGGAAAG 781
```

RESULT 1

LOCUS   AK004136   1510 bp   mRNA   HTC   08-FEB-2001
DEFINITION Mus musculus 18 days embryo cDNA, RIKEN full-length enriched
            library, clone:1110037C17, full insert sequence.
ACCESSION   AK004136
VERSION   AK004136.1 GI:12835201
KEYWORDS   CAP trapper.
SOURCE   Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1110037C17
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2 (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3 (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muranatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4 (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5 (bases 1 to 1510)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,T.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGCGGCCGCCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0. Second strand cDNA
            was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAGGATCCAAGAGCTCAATTAATTAATTAAACCCCCCCCC 3']. cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOLR.
FEATURES   Location/Qualifiers
     source   1..1510
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:106352"
                     /clone="1110037C17"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="18 days embryo"
     CDS   <1..1308
                     /note="putative"
                     /codon_start=1

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	917.4	37.0	1510	192	AK004136	AK004136 Mus muscu
2	721.2	29.1	960	169	BF797306	BF797306 602256926
3	709.6	28.6	761	140	BE795820	BE795820 601590056
4	702	28.3	827	153	BG393395	BG393395 602459323
5	698.2	28.1	762	107	AU124671	AU124671 AU124671
6	689.4	27.8	920	153	BG386090	BG386090 602455264
7	686.4	27.7	712	107	AU126037	AU126037 AU126037
8	679	27.4	823	138	BE619259	BE619259 601473130
9	676.6	27.3	984	139	BE744876	BE744876 601576324
10	676	27.2	676	165	BE260495	BE260495 602455550
11	674.2	27.2	1012	153	BG386348	BG386348 602455550
12	662.6	26.7	812	141	BE867512	BE867512 601443010
13	656.6	26.5	953	169	BE794577	BE794577 602255843
14	655.8	26.4	947	150	BF525432	BF525432 602060517
15	655.6	26.4	735	141	BE902696	BE902696 601677393
16	654	26.4	1124	147	BF342802	BF342802 602015077
17	651.4	26.3	688	140	BE794311	BE794311 602015077
18	649	26.2	949	139	BE744197	BE744197 601577168
19	645.2	26.0	891	175	BG258460	BG258460 602379938
20	641.2	25.9	992	166	BE747163	BE747163 601577254
21	641.2	25.8	936	174	BG116283	BG116283 602318546
22	625.6	25.2	745	141	BE900936	BE900936 601674206
23	624.4	25.2	665	108	AU141334	AU141334 AU141334
24	622.4	25.1	664	166	BE383336	BE383336 601298249
25	621.8	25.1	670	108	AU143668	AU143668 AU143668
26	617.8	24.9	944	172	BF960043	BF960043 602269966
27	616.8	24.9	961	153	BG387104	BG387104 602455813
28	615.4	24.8	754	152	BG324135	BG324135 602430086
29	612.4	24.7	915	147	BF311926	BF311926 601897767
30	610.8	24.6	692	166	BE382253	BE382253 601298656
31	610.4	24.6	790	141	BE892893	BE892893 601435738
32	607	24.5	668	108	AU143074	AU143074 AU143074
33	600	24.2	884	154	BG519751	BG519751 602578902
34	592.2	23.9	696	108	AU138595	AU138595 AU138595
35	592	23.9	677	139	BE742908	BE742908 601574609
36	583.8	23.5	611	154	BG471513	BG471513 602513052
37	582.4	23.5	614	108	AU139902	AU139902 AU139902
38	578.6	23.3	614	152	BG335963	BG335963 602404708
39	578	23.3	940	139	BE743831	BE743831 601577742
40	576.4	23.2	612	166	BE304720	BE304720 601162236
41	556.4	22.4	1497	147	BF315158	BF315158 601899518
42	553.8	22.3	713	136	BE537086	BE537086 601063325
43	553	22.3	553	148	BF434169	BF434169 7099004.x
44	543.4	21.9	938	145	BE260626	BE260626 601145116
45	539	21.7	1112	192	AK003759	AK003759 Mus muscu

ALIGNMENTS

```
117:  gb_est48:*
118:  gb_est49:*
119:  gb_est50:*
120:  gb_est51:*
121:  gb_est52:*
122:  gb_est53:*
123:  gb_est54:*
124:  gb_est55:*
125:  gb_est56:*
126:  gb_est57:*
127:  gb_est58:*
128:  gb_est59:*
129:  gb_est60:*
130:  gb_est61:*
131:  gb_est62:*
132:  gb_est63:*
133:  gb_est64:*
134:  gb_est65:*
135:  gb_est66:*
136:  gb_est75:*
137:  gb_est76:*
138:  gb_est77:*
139:  gb_est78:*
140:  gb_est79:*
141:  gb_est80:*
142:  gb_est81:*
143:  gb_est82:*
144:  gb_est83:*
145:  gb_est84:*
146:  gb_est85:*
147:  gb_est86:*
148:  gb_est87:*
149:  gb_est88:*
150:  gb_est89:*
151:  gb_est90:*
152:  gb_est99:*
153:  gb_est100:*
154:  gb_est101:*
155:  gb_est102:*
156:  gb_est103:*
157:  gb_est104:*
158:  gb_est105:*
159:  gb_est106:*
160:  gb_est67:*
161:  gb_est68:*
162:  gb_est69:*
163:  gb_est70:*
164:  gb_est71:*
165:  gb_est72:*
166:  gb_est73:*
167:  gb_est74:*
168:  gb_est91:*
169:  gb_est92:*
170:  gb_est93:*
171:  gb_est94:*
172:  gb_est95:*
173:  gb_est96:*
174:  gb_est97:*
175:  gb_est98:*
176:  em_esthum29:*
177:  em_esthum30:*
178:  em_esthum31:*
179:  em_esthum32:*
180:  em_esthum33:*
181:  em_estom3:*
182:  em_estpl11:*
183:  em_estro21:*
184:  em_estro22:*
185:  em_estro23:*
186:  em_htc:*
187:  gb_est107:*
188:  gb_est108:*
189:  gb_est109:*

190:  gb_est110:*
191:  gb_est111:*
192:  gb_htc:*
193:  em_gss_fun:*
194:  em_gss_fun1:*
195:  em_gss_hum1:*
196:  em_gss_hum2:*
197:  em_gss_hum3:*
198:  em_gss_hum4:*
199:  em_gss_hum5:*
200:  em_gss_hum6:*
201:  em_gss_hum7:*
202:  em_gss_hum8:*
203:  em_gss_inv1:*
204:  em_gss_inv2:*
205:  em_gss_inv3:*
206:  em_gss_other:*
207:  em_gss_pln1:*
208:  em_gss_pln2:*
209:  em_gss_pro:*
210:  em_gss_rod1:*
211:  em_gss_rod2:*
212:  em_gss_rod3:*
213:  em_gss_rod4:*
214:  em_gss_rod5:*
215:  em_gss_vrt1:*
216:  em_gss_vrt2:*
217:  em_gss_vrt3:*
218:  gb_gss1:*
219:  gb_gss2:*
220:  gb_gss3:*
221:  gb_gss4:*
222:  gb_gss5:*
223:  gb_gss6:*
224:  gb_gss7:*
225:  gb_gss8:*
226:  gb_gss9:*
227:  gb_gss10:*
228:  gb_gss11:*
229:  gb_gss12:*
230:  gb_gss13:*
231:  gb_gss14:*
232:  gb_gss15:*
233:  gb_gss16:*
234:  gb_gss17:*
235:  gb_gss18:*
236:  gb_gss19:*
237:  gb_gss20:*
238:  gb_gss21:*
239:  gb_gss22:*
240:  gb_gss23:*
241:  gb_gss24:*
242:  gb_gss25:*
243:  gb_gss26:*
244:  gb_gss27:*
245:  gb_gss28:*
246:  gb_gss29:*
247:  gb_gss30:*
248:  gb_gss31:*
249:  gb_gss32:*
250:  gb_gss33:*
251:  gb_gss34:*
252:  em_gss_inv4:*
253:  em_gss_rod6:*
254:  em_gss_rod7:*
255:  em_gss_rod8:*
256:  gb_gss35:*
257:  gb_gss36:*
258:  gb_gss37:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 16:52:01 ; Search time 1834.93 seconds
                                               (without alignments)
                                               12781.158 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgcttgctcgct.........agaaggtcaagagccagtga 2481
Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0
Searched: 1028815 seqs, 472642675Q residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Database :   EST:*

  1: gb_est1:*
  2: gb_est2:*
  3: gb_est3:*
  4: gb_est4:*
  5: gb_est5:*
  6: gb_est6:*
  7: gb_est7:*
  8: gb_est8:*
  9: gb_est9:*
 10: gb_est10:*
 11: gb_est11:*
 12: gb_est12:*
 13: gb_est13:*
 14: gb_est14:*
 15: gb_est15:*
 16: gb_est16:*
 17: gb_est17:*
 18: gb_est18:*
 19: gb_est19:*
 20: gb_est20:*
 21: gb_est21:*
 22: gb_est22:*
 23: gb_est23:*
 24: gb_est24:*
 25: gb_est33:*
 26: gb_est34:*
 27: gb_est35:*
 28: gb_est36:*
 29: gb_est37:*
 30: gb_est38:*
 31: gb_est39:*
 32: gb_est40:*
 33: em_estba:*
 34: em_estfun:*
 35: em_esthum1:*
 36: em_esthum2:*
 37: em_esthum3:*
 38: em_esthum4:*
 39: em_esthum5:*
 40: em_esthum6:*
 41: em_esthum7:*
 42: em_esthum8:*
 43: em_esthum9:*
 44: em_esthum10:*
 45: em_esthum11:*
 46: em_esthum12:*
 47: em_esthum13:*
 48: em_esthum14:*
 49: em_esthum15:*
 50: em_esthum16:*
 51: em_esthum17:*
 52: em_esthum18:*
 53: em_esthum19:*
 54: em_esthum20:*
 55: em_esthum21:*
 56: em_esthum22:*
 57: em_esthum23:*
 58: em_esthum24:*
 59: em_esthum25:*
 60: em_esthum26:*
 61: em_esthum27:*
 62: em_esthum28:*
 63: em_estin1:*
 64: em_estin2:*
 65: em_estin3:*
 66: em_estin4:*
 67: em_estin5:*
 68: em_estom1:*
 69: em_estom2:*
 70: em_estov1:*
 71: em_estov2:*
 72: em_estpl1:*
 73: em_estpl2:*
 74: em_estpl3:*
 75: em_estpl4:*
 76: em_estpl5:*
 77: em_estpl6:*
 78: em_estpl7:*
 79: em_estpl8:*
 80: em_estpl9:*
 81: em_estpl10:*
 82: em_estro1:*
 83: em_estro2:*
 84: em_estro3:*
 85: em_estro4:*
 86: em_estro5:*
 87: em_estro6:*
 88: em_estro7:*
 89: em_estro8:*
 90: em_estro9:*
 91: em_estro10:*
 92: em_estro11:*
 93: em_estro12:*
 94: em_estro13:*
 95: em_estro14:*
 96: em_estro15:*
 97: em_estro16:*
 98: em_estro17:*
 99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

REFERENCE   1 (bases 1 to 762)
AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
TITLE   HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES   Location/Qualifiers
     source   1..762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RM4000375"
                     /clone_lib="NT2RM4"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
BASE COUNT   180 a   212 c   165 t   3 others
ORIGIN
                                                         202 g

Query Match   28.1%;   Score 698.2; DB 107; Length 762;
Best Local Similarity   97.9%;   Pred. No. 5.8e-172;
Matches 748; Conservative   0; Mismatches   11; Indels   5; Gaps 4;

Qy   852 cactcatgaaggaagagagagattttggctgaagacggctgttgttactcctccagaatcctggtgc 911
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CACTCATGAAGGAAGAGAGAGATTTTGGCTGAAGACGGCTGTGTTACTCCTCCAGATCCTGGTGC 60

Qy   912 tgtctttggttggtggtagaatgtccagatgaaagcttcattcattcaaccatctgtgagaatgc 971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGCTTTTGTGGTGGTGGTAGAAATGTCCAGATGAAAGCTTCATTCAACCATCTGTGAGAATGC 120

Qy   972 cacctttcagaggtaccaaggaaagcagatgcccccctggcctttggtttggttcacaatggc 1031
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACCTTTCAGAGGTACCAAAGGAAAGCAGAATGCCCCCGTGGCCTTTGGTGGTTCACATGGC 180

Qy   1032 ccagcatctgtcttgtgacagcaggtaccagcagtggatggaaggtttgggcctga 1091
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CCCAGCACTCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAAGGTTTGGGCCTGA 240

Qy   1092 cacccagcacttgtcctgaatgagaacttgtgctcctcagttcacaccttcgcaccacaa 1151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCACAACCTTCGCACCACAA 300

Qy   1152 gattcaaaacccagctcaacctcatccaccccggacatcttccccccctgctcaccagttccg 1211
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GATTCAAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCG 360

Qy   1212 ctgtaagagaggaggcccccacccctcagtgttccaccatggttcaggtgaatgcctcctcaa 1271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTGTAAGAGAGGAGGGCCCCACCCCTCAGTGTGCCACCATGGTTCAGGGTGAATGCCCTCCTCAA 420

Qy   1272 gtaccagctccgtcccagcgagggagtggcagaggatgccatttacttgtgcaatcctga 1331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTACCAGCTCCGTCCCAGCGAGGGAGTGGCAGAGGGATGCCATTAATTACTTGCAATCCTGA 480

Qy   1332 ggaattcaatagttgaggcgctgcaggcagcttcccaaacttcccagcggatgaaactacag 1391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GGAATTCAATAGTTGAGGCGGCTGCAGCTTCCCAACTTCCATCAGAGCCCGTGCAGGAGTACAG 540

Qy   1392 gaggagtgcgcaggacgcgccccagcccccagccagcagcagagaagaagaagtcagtaccagaaatcat 1451
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAGGAGTGCGCAGGACGCGCCCCAGCCCCAGCCAGCAGCAGAGAAAAGAAGAGTCAGTACCAGAAATCAT 600

Qy   1452 cttccttggaacaggggtcctgccatcccgatgaagattcgaaatcgtcaagtgccacacttgt 1511
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CTTCCTTGGACAGGGGTCTGCCCATCCCGATGAAGAATTCGAAATGTCAGTTGTT 659

Qy   1512 caacataaagccccgacacacgtctctgtac-tggactgtggtgaggcacattgggcagc 1570
              ||||| ||| ||||||| ||||||||||||| || |||||||||| |||||||||||||||
Db   660 CAACATAAGCCCCGACACGTCTCTGCTACTTGGACTGTGGTGANGGCACATTTGG--CAC 717

Qy   1571 tgtgccgtcattacggaagaccaggtgacagggtcctgggcacc 1614
              |||| || ||||||||||||||||||||||||||||||||||||
Db   718 TTNGCCCGCATTTCNGAGACCAAGTGG-CAGGGTCCTGGGCACC 760

RESULT 6
BG386090

LOCUS   BG386090   920 bp   mRNA   EST   12-MAR-2001
DEFINITION 602455264F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583679 5',
            mRNA sequence.
ACCESSION   BG386090
VERSION   BG386090.1 GI:13279536
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 920)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM1307 row: p column: 16
            High quality sequence stop: 736.
FEATURES   Location/Qualifiers
     source   1..920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583679"
                     /clone_lib="NIH_MGC_15"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
BASE COUNT   196 a   261 c   272 g   190 t   1 others
ORIGIN

Query Match   27.8%;   Score 689.4; DB 153; Length 920;
Best Local Similarity   91.9%;   Pred. No. 1.2e-169;
Matches 786; Conservative   0; Mismatches   56; Indels   13; Gaps 5;

Qy   896 ctccagaatcctggtgctgctcttttgtggttggtagaaatgtccagatgaaagcttcattcaac 955
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 CTCCAGATCCTGGTGCTGCTGC-TTTGTGGTGGTGGTAGAAATGTCCAGATGAAAGCTTCATTCAAC 60

```
RESULT  7

LOCUS       AU126037     712 bp    mRNA            EST       23-OCT-2000
DEFINITION  AU126037 NT2RM4 Homo sapiens cDNA clone NT2RM4002610 5', mRNA
            sequence.
ACCESSION   AU126037
VERSION     AU126037.1  GI:10950753
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3931
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Helix
            Research Institute; cDNA to mRNA: Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT2RM4"
                     /clone="NT2RM4002610"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
BASE COUNT      173 a    198 c    204 g    134 t      3 others
ORIGIN

Query Match        27.7%;   Score 686.4;  DB 107;  Length 712;
Best Local Similarity  99.4%;   Pred. No. 7e-169;
Matches  687;  Conservative  0;  Mismatches  4;  Indels  0;  Gaps  0;
```

```
Qy      1 atgtgggggcgcttgctgctgctcgcgtcgcgcggcggacgacgcaccaatgtcgcaggagc    60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 ATGTGGGGGCGCGCTTGCTGCTGCTCGCGTCGCGCGGCGGACGACGCACCAATGTCGCAGGAGC    81

Qy     61 accatatcgcaggacaccgccgcccgccgccagccgccgcaagaccccgtgcgcgcacctg   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 ACCATATCGCAGGACACCGCCGCCCGCCGCCAGCCGCCGCAAGACCCCGTGCGCGCACCTG   141

Qy    121 cgcacgcgagaagacgcggaccgtcgggggtgctccgagcgcggcccaaacaccgtgtacctg   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 CGCACGCGAGAAGACGCGGACCGTCGGGGGTGCTCCGAGCGCGGCCCAAACACCGTGTACCTG   201

Qy    181 caggttgggcaggggtagcccgggactcgggggtcgcgctctaacgtctctccgagttc     240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 CAGGTTGGGCAGGGGTAGCCCGGGACTCGGGGGTCGCGCTCTACGTCTTCTCCGAGTTC     261

Qy    241 aaccggtatctctcaactgtggagaaggcgttcagagactcatgcaggagacacaagtta    300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 AACCGGTATCTCTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGACACAAGTTA    321

Qy    301 aaggttgctcgcctggacacacatatctcctgacacgaatgcactggtctaatgtggggc   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 AAGGTTGCTCGCCTGGACACACATATTCCTGACACAACATATTCCTACACGAATGCACTGGTCTAATGTTGGGGC    381

Qy    361 ttaagtggaatgattcttacttaaaggaaaaccgggcttccaaagtgtactttctgga    420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 TTAAGTGGAATGATTCTTACTTAAAGGAAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA    441

Qy    421 ccttccacaactgaaaaatatacctcgaagcaatcaaaaatattttggttccattgaaagga    480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 CCTTCCACAACTGGAAAAATACCTCGAAGCAATCAAAAATATTTTCGTTCCATTGAAAGGA    501

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt    540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAAACCATGACAGTT    561

Qy    541 taccagatcccatacacacagtgaacagaggaggggaagcacacaacctgggcagagtcca    600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    562 TACCAGATCCCCATACACACAGTGAACAGAGGAGGGGAAGCACCAACATGGCAGAGTCCA    621
```

```
Qy    956 ccatctgtgagaatgccaccttcagaggtaccaagagaaagccagatgcccccgtgtggct   1015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGAGAAAGGGCAGAATGCCCCCGTGGGCT    120

Qy   1016 tggtggttcacatgtggcccagacatctgtgcttgtgtggacagcaggtaccagcagtggatg   1075
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGTGGTTCACATGGCCCCAGACATCTGTGCTTGTGTGGACAGCAGGTACCAGCAGTGGATGG    180

Qy   1076 agaggtttgggcctgacacccagacacttggtcctgaatgatgagaactgttgcctcagttcaca   1135
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGAGGTTTTGGGCCTGACACCCAGACACTTGGTCCTGAATGATGAGAACTGTTGCCTCAGTTCACA    240

Qy   1136 accttcgcagccacaagattcaaacaccagctcaacctccacaccggaacactttcccc   1195
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACCTTCGCAGCCACACAAGAATTCAAACACCAGCTCAACCTCATCCACCCAGTTCACA    300

Qy   1196 tgctcaccagttccgctgtaaggaggggcccccacccttcagtgtcgtgccatgttcagg   1255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGCTCACCAGTTTCCGCTGTAAGGAGGGGCCCCCACCCTTCAGTGTCGTGCCATGGTTCAGG    360

Qy   1256 gtgaatgcctctcaagtaccagctccgtccagggaggggagtggtgatgccatta   1315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTGAATGCCTCTCAAGTACCAGCTCCGTCCCAGGGAGGGGAGTGGCAGAGGGAATGGCCATTA    420

Qy   1316 ttacttgcaatcctgaggaattcaatagttgaggcgctgacgcttcccaacttccagcaga   1375
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGACGCTTCCCAACTTCCAGCAGA    480

Qy   1376 gcgtgcaggtagtacaggaggaggtgcggcaggacggcccagcccagcagagagaaaagaagtc   1435
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCGTGCAGGAGTACAGGAGGAGGTGCGGCAGGACGGCCCAGCCCCAGCAGAGAGAAAAGAAGTC    540

Qy   1436 agtaccagagaatcatctcttcctggaacagggtctgccatccccgatgaagattcgaaatg   1495
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGTACCCAGAGAAATTCATCTTCCTTGGAACAGGGTCTG-CATCCCGATGAAGAATTCGAAATG    599

Qy   1496 tcaagtgccacacttgtcaacataaggccccgacaacgtctctgttactgggactcgttggtcagg   1555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 TCAGTGCCACACTTGTCAACATTAAGGCCCCGACACGTCTCTGCTTACTGGACTGTGGTCAGG    659

Qy   1556 gacacattgggcagctgtgccgctcattacggtgacaggtggacaggtgtcctgggcacc   1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 GCACATTTGGGCAGCTGTGTGCCGTCATTACGGAGACCAGGTGGCC---GTCCTGGGCACCC    717

Qy   1616 tggctgctgtttgtgtcccacctgcacgcagatcaccacacggggcttgccaagtatct   1675
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    718 TGG-----TGCGTGTTGTGTCCCACTGCAGCAGAATTACACCGGGTG-----CAAGTATCT    768

Qy   1676 tgctgaagagaaacgagcctggacatcttttgggaaagccgcttcacccttgctggtgg   1735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    769 GGCTGCAGAGAAACGAGCCTGGTGTCTGGGAAGGCGTTCACCTTGTCCCACAGTAAGGGG    828

Qy   1736 ttgcccccaaacagc   1750
          ||  |  |||  ||
Db    829 TCGGGGGCCACCGGC    843
```

```
Qy    601 gaaaggccttcaacctgccaagctcaagttccagagagcgatcttcagactccgagtcggaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622 GAAAGGCCTTCTCAGCAAGGCTCAGTCCAGAGCGGATCTTCANACTCNCGAGTCGGAATGAAAAT 681

Qy    661 gagccacaccttccaacatggtgttagccaga 691
          ||||||||||||||||||||||||||||||
Db    682 GAGCCACACCTTTCACATGGTGNTAGCCAGA 712
```

RESULT 8
LOCUS   BE619259   823 bp   mRNA   EST   20-OCT-2000
DEFINITION   601473130F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5',
             mRNA sequence.
ACCESSION   BE619259
VERSION   BE619259.1 GI:9850197
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 823)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9636 row: k column: 08
             High quality sequence stop: 695.
FEATURES   Location/Qualifiers
     source   1..823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3876223"
                     /tissue_type="large cell carcinoma"
                     /clone_lib="NIH_MGC_68"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."
BASE COUNT   185 a   219 c   160 t   259 g
ORIGIN

Query Match   27.4%; Score 679; DB 138; Length 823;
Best Local Similarity   92.7%; Pred. No. 6.4e-167;
Matches 757; Conservative   0; Mismatches   55; Indels   5; Gaps   4;

```
Qy   1562 ttgggcagctgtgccgtcattacggagaccaggtggacaggtcctggggacaccctggctg 1621
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 TTGGGCACGNGCTGCCGTCATTACGGAGACAACCAGGTGGACAGGTGTCTGGGCACCCTGGGCTA 69

Qy   1622 ctgtgtttgtgtcccacctgcacgcagatcaccacacggcttgccaagtatcttgctgc 1681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 CTGTG-TTGTGTCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGC 128

Qy   1682 agagagaaacgcgccttggcatcttggaaagccgcttcaccctgctgttggttgcc 1741
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 AGAGAGAAACGCCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCC 188

Qy   1742 ccaaccaagctcaaagctgggctccaaggagagtcctgcacc 1801
          ||||||||||||||||||||||||||||||||||||||||
Db    189 CCAACCAGCTCAAAGCCTGGCTCCAGCAGTGACCACACAACCAGTGCCAGGAGGTCCTGCCACC 248

Qy   1802 acatcaagtatgattcctgccaaatgcctcaggaagggctgagatctccagtcctgcag 1861
```

RESULT 9
LOCUS   BE744876   984 bp   mRNA   EST   15-SEP-2000
DEFINITION   601576324F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3837533 5',
             mRNA sequence.
ACCESSION   BE744876
VERSION   BE744876.1 GI:10158868
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 984)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM522 row: o column: 06
             High quality sequence stop: 726.
FEATURES   Location/Qualifiers
     source   1..984
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3837533"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"

```
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 ACATCAGTNATGNATTCCTCCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAG 308

Qy   1862 tggaaagaattgatcagttcgtctgttcggaacatgtgatttgatttggaaagagttcagcacgtgtc 1921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 TGGAAAGAATTGATCAGTTCGTCTGTTGGAACATGTGATTTGGAAAGAGTTTCAGACCTGTC 368

Qy   1922 tggtcggcactgcaagcatgcgttctggctgtcgcgctggtgcacacctctggctggaaag 1981
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 TGGTGCGGCACTGCAAGCACATGCGTTTGGCGTGTCGCGTGTGCACACTCTGGCTGGAAAG 428

Qy   1982 tggtctattccgggacaccatgcccctgcctgcgagctcttggtccggatggggagaatgcca 2041
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 TGGTCTATTCCGGGACACCATGCCCCTGCCTGCGAGCTCTTGGTCCGGATGGGGAGAATGCCA 488

Qy   2042 ccctcctgatacatgaagcacacctgggaagtatcggaaggaaggacagtggaaaagaaga 2101
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 CCCTTCCTGATACATGACATGAAGCCACACCCTGAAGGAAGATGGTTGGAAGGAAGACAG-GGAAAAGA 547

Qy   2102 cacacagcaacaaagtacccaaagccaatccagcgtggggatgcgggatgaacgcggagttcatta 2161
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    548 CACACAGCAACAAAGTCCCAAAGCCCATCAGCGTGGGGATGAACGCGGAGTTCATTA 607

Qy   2162 tgctgaaccacctcagccagcgctatgccaaggtccccctccttcagccccaaacttcagcg 2221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    608 TGCTGAACCACTTCAGCCAGCGGTATGCCAAGGTCCCCAGTTCCTTCAGCCCCAAAACTTCAGCG 667

Qy   2222 agaaagtgggagttgcctttgaccacatgaaaggtctgctttggagagacttccaacaatgc 2281
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    668 AGAAAGTGGGAGTTGCC--TTGACCACATGAAAGGTCTGC-TTGGGGACTTTCCAACAAANTG 724

Qy   2282 ccaagctgattccccactgaaagccccctgtttgctggtggggacatcgaggagagatggaggagc 2341
          ||||||||||| |||  ||||||||||||||||||||||||||||||||||||||
Db    725 CCAAGGTGGATTCCCCAATGAAAGCCTGTGTGTCGGGGACACGAGGGCCAAGGGGGGGGGGGCC 784

Qy   2342 gcagggagaaagcggagctgcgcggcggggcggc 2378
          ||| |||||  |||  |||||||||||||||||||
Db    785 GGAAACGGGAGCTGGGAAAAGGGGGGGGGGGGGGGGGCCCCCTGGGGC 821
```

RESULT 10

LOCUS   BE260495   676 bp   mRNA   EST   26-OCT-2000
DEFINITION 6011507702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
            mRNA sequence.
ACCESSION   BE260495
VERSION   BE260495.1 GI:9131807
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 676)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM176 row: d column: 01
            High quality sequence stop: 672.
FEATURES   Location/Qualifiers
     source   1..676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503184"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   154 a   207 c   176 g   139 t
ORIGIN

Query Match   27.2%; Score 676; DB 165; Length 676;
  Best Local Similarity   100.0%; Pred. No. 3.7e-166;
  Matches 676; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1093 accagcacttgtcctgaatgagaactgtgcctcagttcacaacttcgcagccacaag 1152
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACCCAGCACTTGGTGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCACACAAG 60

Qy   1153 attcaaacccagctcaaactcatccaacccggacatcttcccctgctcaccagtttccgc 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTCAAACCCAGCTCAAACTCATCCAACCCGGACATCTTCCCCTGCTCACCAGTTTCCGC 120

Qy   1213 tgtaagaacgagggcccaccctcagtgttgtgcccatggttcaggtgatgcctcctcaag 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGTAAGAACGAGGGCCCACCCTCAGTGTGCCCATGGTTCAGGTGATGCCTCCTCAAG 180

Qy   1273 taccagctccgtcccagagggagtggcagagggggtggccattattacttgcaatcctgag 1332
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGGGTGGCCATTATTACTTGCAATCCTGAG 240

Qy   1333 gaattcatagttgaggcgctgcagcttcccaacttccagcagaggcggtgcaggagtacg 1392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGGCGGTGCAGGAGTACG 300

Qy   1393 aggagtgcgcaggacggacggcgcccagcccagcccagcagagaaaagaagtcagtacc 1452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGGAGTGCGCAGGACGGACGGCGCCCAGCCCAGCCCAGCAGAGAAAAGAAGTCAGTACC 360

Qy   1453 ttcctggaacaggtctcgccatcccgatgaagattcccagaaatgtcagtgccacacttgtc 1512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ...

BASE COUNT   227 a   259 c   316 g   182 t
ORIGIN

                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

Query Match   27.3%; Score 676.6; DB 139; Length 984;
  Best Local Similarity   98.2%; Pred. No. 2.9e-166;
  Matches 726; Conservative   0; Mismatches   9; Indels   4; Gaps   4;

Qy   682 gttagccagaagaagaggggtcaggagactcctccctgctgctagcttcatctgtaagctt 741
              ||||||||||| |||||| |||||||||| ||| ||||| ||| ||||| |||||||||||||
Db   12 GTTAGCCAGAGAAGAGAGGGGTCAGGGACTTCTCCTGTCGTCAGAGCTTTCATCTGTAAGCTT 71

Qy   742 cacttaaagaagagaggaaacttcttggtgctcaaagcaaaggagggcctcccagttggg 801
              ||||||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||
Db   72 CACTTAAAGAGAGGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGGATGGGCCTCCCAGTTGGG 131

Qy   802 acagctgccatcgtccatcattgctgctgtcaaggacgggaaagcgatcactcactcatgaa 861
              ||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||||
Db   132 ACAGCTGCCATCGTCCATCATTGCTGCTGTCAAGGACGGGAAAAGCGATCACTCATGAA 191

Qy   862 ggaaagagagattttggctgaagagctgtgtactcctccagatcctgtgtgctgctttgtg 921
              ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   192 GGAAGAGAGA-TTTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGTGTGCTGC-TTTGTG 249

Qy   922 gtggtagaaatgtccaagatcattcaatccacccacctttcag 981
              ||||||||||||||||| || ||| |||||||||| |||| |||
Db   250 GTGGTAGAAATGTCCAGATGAAAGCTTCAATCCACCCATCTGTGAGAATGCCACCTTTCAG 309

Qy   982 aggtaccaaggcagaatggcccccgtggccttggtggttcacatggccccagcatct 1041
              ||||||||||||||||| ||||| |||||||||||||||||||||||||||||||||
Db   310 AGGTACCAAGGCAGAAAGGCCAGAATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCT 369

Qy   1042 gtgcttgtgggacaagcagcagtggatgaggtttggcctgacacccacgac 1101
              |||||||||||||||||||||||||| |||| ||||||| ||||||||||||
Db   370 GTGCTTGTGGGACAAGCAGCAGTGGATGGATGGAGGATGGGCCTGAC-TTGGGCCTTGGCCTGACACCCACGAC 428

Qy   1102 tcggtcctgaatgagaactgtgcctcagttcacaacttcgcagccacaag 1161
              | ||||||||||||||||||||||||||||||||||| |||||| |||||
Db   429 TTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCACACAAG 488

Qy   1402 caggacggccagcccag 1420
              ||||| |||||||||||
Db   728 CAGGAGGACCAGCCCAGCCCAG 746

Qy   1342 gttgaggcgctgcagcttcccaacttccagcagagcctgcaggaggaggtgcg 1401
              |||||||||||| ||||||||||||||||||||||| |||||||||||||||||
Db   669 GTTGAGGCGCTG-AGCTTCCCAACTTCCAGCAGAGCCTGCAGGAGGAGGTGCG 727

Qy   1282 cgtcccaaggaggaggatggccaagagggatgcagagggatgacattattacttgcaatcctgag 1341
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   609 CGTCCCAAGGAGGAGGATGGCCATTATTACTTGCAATCCTGAGGAATTCATA 668

Qy   1222 gaggcgcccaccctcagtgttgccatggttcaggtgatgcctcctcaagtaccagctc 1281
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   549 GAGGCGCCCACCCTCAGTGTGCCCATGGTTCAGGTGATGCCTCCTCAAGTACCAGCTC 608

Qy   1162 cagctcaaactcatccaacccggacatcttcccctgctcaccagtttccgc 1221
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   489 CAGCTCAAACTCATCCAACCCGGACATCTTCCCCTGCTCACCAGTTTCCGC 548

Qy   1222 gaggcgcccaccctcagtgtt... 1281
Db   549 ...

RESULT 11

```
Db    361 TTCCTTGGACAGGGTCTGCCATCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTC 420
                                                              
Qy   1513 aacataagccccgaacgtctctgtactggactgtgtggtgaaggggacacatttgggcagctg 1572
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACATAAGCCCCGAACACACGTGCTCTCTACTGGACTGTGGTGAGGGCACAATTGGGCAGCTG 480
                                                              
Qy   1573 tgccgtcattaccggagaccaaggtggacagggtgcaccctgggctgctgttgtttgtg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCCGTCATTACCGGAGACCAAGGTGGACAGGGTCCCTGGGCACCCTGGCTGCTGTTTGTG 540
                                                              
Qy   1633 tcccaacctgcacgcagatcacccaacgggcttgccaagtatcttgctgcagagaacgc 1692
          |||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCCACCTGCACGCAGATCACCACACGGGCCTTGCCAAGTATCTTTGCTGCTGCAGAGAAACGC 600
                                                              
Qy   1693 gccttggcatcttttgggaaagccgcttcacccttgctggttgtcccccaaccagctc 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCTTGGCATCTTTTGGGAAAAGCCCTTCACCCTTTGCTGGTTGTCCCCAACCAGCTC 660
                                                              
Qy   1753 aaagcccctggcttccagc 1768
          ||||||||||||||||||
Db    661 AAAGCCTGGCTCCAGC 676
```

Query Match   27.2%;   Score 674.2;   DB 153;   Length 1012;
Best Local Similarity   96.2%;   Pred. No. 1.2e-165;
Matches 743;   Conservative   0;   Mismatches   24;   Indels   5;   Gaps   5;

RESULT 12

```
Qy    896 ctccagatctcctggtgctgctttgttcggtgtgttagaatgtccagatgaaagcttcattcaac 955
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CTCCAGATCCTCGGTGCTGCTTTTGTGTGGTGTTAGAATGTCCAGATGAAAGCTTCATTCAAC 61
                                                              
Qy    956 ccatctgtgagaatgccaccttcagaggtaccaaggaaggcagagatgcccccgtggcct 1015
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CCATCTGTGAGAATGCCACCTTCAGAGGTACCAAGGAAAGCCCCCGTGGCCT 121
                                                              
Qy   1016 tggtgttcacatgggccccagcatctgtgcttgttggtgacagcagcaagtggatgg 1075
          |||||||||||||||||||||||||||||||||||||||||||||
Db    122 TGGTGGTTCACATGGGCCCCAGCATCTGTGCTTGTTGGTGGACAGCAGCAGTGGATGG 181
                                                              
Qy   1076 agaggtttggggcctgacaccccag-cacttggtcctgaatgaggaactgtgcctcagttcac 1134
          |||||||||||||||||||||||  |||||||||||||||||||||||||
Db    182 AGAGGTTTGGGGCCTGACACCCCAGTCACTTGGTCCTGAATGAGGAACTGTGCCTCAGTTCAC 241
                                                              
Qy   1135 aaccttcgcagccacaagattcaaaaccccagctcaaactcatcaccccggacatcttccc 1194
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    242 AACCTTCGCAGCCACAAGATTCAAAACCCCAGCTCAAACTCATCACCCCGGACATCTTCCC 301
                                                              
Qy   1195 ctgctcaccagtttccgctgtaagaagaggaggggccccaccccctcaagtgtgccatggttcag 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 CTGCTCACCAGTTTCCCCTGTAAGAAGGAGGGGCCCCACCCCTCAGTGTGCCATGGTTCAG 361
                                                              
Qy   1255 ggtgaatgcctcctccaagtaccagctcccgtcccaagaggggagggggatgccatt 1314
          ||||||||||||||||||||||||||||||||||||||||||
Db    362 GGTGAATGCCTCCTCCAAGTACCAGCTCCCGTCCCAGGAGGGGAGGGGGATGCCATT 421
                                                              
Qy   1315 attacttgcaatcctgaggaattcatagttgaggcggctgcagcttccaacttccagcag 1374
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGGCTGCAGCTTCCAACTTCCAGCAG 481
                                                              
Qy   1375 agcgtgcaggagtacagggaggagtcgcggacacgtctctgctactggtggtga 1434
          |||||||||||||||||||||||||||||||||||||||||
Db    482 AGCGTGCAGGAGTACAGGGAGTGCGCCAGCACGTCTCTGCTACTGGTGGTGA 541
                                                              
Qy   1435 cagtaccccagaaaatcatctccttggaacacaggtctgccatccccgatgaaagattcg-aaa 1493
          ||||||||||||||||||||||||||||||||||||||||||  |||
Db    542 CAGTACCCAGAAAATCATCTCCTTGGAACACAGGGTCTGCCATCCCCGATGAAAGATTCGAAAA 601
                                                              
Qy   1494 tgtgtcagtgccacactgtcacacataagccccgaacacgtctctgctactggtggtga 1553
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    602 TGTCAGTGCCACACTTGTCACACATAAG-CCCGAACACGTCTCTGCTACTGGTGGTGA 660
                                                              
Qy   1554 gggcacacattgggcagctgtgccgtcattac-ggagaccaaggtgacaggtgtcctggca 1612
          |||||||||||||||||||||||||  |||||||||||||||||||||
Db    661 GGGCACAATTGGGCAGCTGTGCCGGTCATTTACCGGAGAACCACTGGGACGGGGTCCTGGGC 720
                                                              
Qy   1613 ccctggcgctgtgtttgtgt-cccaccctgcacgcagatcaccacacggggct 1663
          |||| |||||||||||||  ||||| ||| ||||||||||||||||
Db    721 ACCTGNGTTGTGTGTTTGTGTGTCCCCACCTTGCACGCCGATCAACACAAACGGGGCT 772
```

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9561 row: c column: 03

FEATURES
source
    Location/Qualifiers
    1..812
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3847226"
    /clone_lib="NIH_MGC_65"
    /tissue_type="adenocarcinoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
    Average insert size 1.8 kb. Library constructed by Life
    Technologies."
    High quality sequence stop: 686.

BASE COUNT   200 a   216 c   220 g   176 t
ORIGIN

Query Match   26.7%;   Score 662.6;   DB 141;   Length 812;
Best Local Similarity   97.4%;
Matches 716; Conservative 0; Pred. No. 1.3e-162; Mismatches 14; Indels 5; Gaps 4;

```
Qy   724 gctttcatctgtaagcttcacttaaagaagaagaggaaacttcttggtgtctcaaagcaaaggag 783
         |||||||||||||||||||||||| ||||||||| ||||||||||||| |||| |||| ||||||
Db     1 GCTTTCATCTGTAAGCTTCACTTAAAGAGAGAAGAAAACTTCTTTGGTGCTGTCAAAACAAAGGAG 60

Qy   784 atgggcctcccaagttgggacagtgccagcgtcccatcgctccccatcattgctctgtcaaggacggg 843
         ||||||||| |||||||||||| |||||| |||||||||||| ||| ||||||||||||||||||||||
Db    61 ATGGGCCTCCCAGTTGGGACAGCTGCCATCCGTCCCATCATTGCTCTGTCGTGTCAAGGACGGG 120

Qy   844 aaaagcatcatcagaaagaagagaagagatttggctgaagagagctgtgtactcctccagat 903
         |||||||||||||||||||||||||||||||| || ||||| ||||||||||||||||||
Db   121 AAAAGCATCACTCATGAAGGAAGAGAGAGATTTTGGCTGAAGAGAGCTGTGTACTCCTCCAGAT 180

Qy   904 cctggttgctgctttgttcggtcgtagaaatgtccagatgaaagcttcattcaaccatcatgt 963
         |||||||| |||||| |||||||||| |||| ||||| |||||||||||| ||||||||||||
Db   181 CCTGGTCCTGC·TTTGTGGTGTGTAGAATGTCCAGATGAAAAGCTTCATTCAACCCATCTGT 239

Qy   964 gagaatggccaccttcagagaggtaccaagaagagatgccgccccgtggcccttggttggtt 1023
         |||||||||||||||||||||| |||||||||||||| ||||||||||| ||||||||||||||
Db   240 GAGAATGCCACCTTTCAGAGGTACCAAGGACAGATGCCCGCCCCGTGGCCTTGGTTGGTT 299

Qy  1024 cacatggcccagcacactcgtgcgttgtggaacgcagtgacacagcggatggagaggtt 1083
         |||||||||||||||||| ||||||| |||||||||||||||||||||||||||||||||
Db   300 CACATGGCCCCAGCACATCTGTGCGTTGTGGGACAGCGGTACCACGGTGGAGAGGTTT 359

Qy  1084 gggcctgacaccagcacttcggtccctgaatgaagaactgtgctcagttcacaaccttcgc 1143
         ||||||||||| |||||||||||||||||| || |||| ||||||||||||||||||||||
Db   360 GGGCCTGACACCAGCACTTGGTCCCTGAATGAGAACTGTGCTCTCAGTTCACAACCTTCGC 419

Qy  1144 agccacacaagattcaaacccagctcaacctcatccaaccggacaatctcttcccccctgctcacc 1203
         ||||||| |||||||||||||||||| |||| ||||||||||||||||| |||||||||| ||||||
Db   420 AGCCACAAGATTCAAACCCAGCTCAACCTCATCCAACCGGACAATCTTCCCCGGTGCTCACC 479

Qy  1204 agtttccgctgtaaagaaagaggaggggccccaccacctcagtgtgtgccatggttcaggggtgaatgc 1263
         |||||| ||||||||||||||||||||||||| ||||||||||| |||||||||||||| |||| ||||
Db   480 AGTTTCCGCTGTAAAGAAAGAGGAGGGGCCCCACCACCTCAGTGTGTTCAGGTTCAGGGTGAATGC 539

Qy  1264 ctcctcaagtaccagctccgtcccaggaggg··agttgcagagggatgccattattact· 1320
         |||||||| |||||||||||||||||||||| |||||||||| |||| ||||| |||||
Db   540 CTCTCAAGTACCAGCTCCGTCCCAGGAGGGTCCAGTGGGCAGTGGATGCCATTATTACTC 599

Qy  1321 tgcaatcctgaggaattcatagttgagggcgctgagctcttccaaggcacagagcgtg 1380
         ||||| ||||||||| ||||||||| ||||| ||||||||||||||||||||| ||||
Db   600 TGCAATCCTGAGGAATTCATAGTTAGTTGAGGCGGTTTCCCAAGCACAGAGCGTG 658

Qy  1381 caggagtacagaggaagtgcgcaggacggcccagccagcgcccagaagaaagtcagtac 1440
         |||||||||||||||||| |||||||| |||||||||||||||||| |||||||||||||
Db    ... caggagtacagaggaagtgcgcaggacggcccagccagcgcccagagagaaagtcagtac 1440
```

RESULT 13
BF794577
LOCUS   BF794577   953 bp   mRNA   EST   12-JAN-2001
DEFINITION   602258043F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4339187 5',
             mRNA sequence.
ACCESSION   BF794577
VERSION   BF794577.1 GI:12099631
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 953)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Louis Staudt, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9949 row: e column: 12
             /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally. Oligo-dT primed.
             Average insert size 1.867 kb. Library enriched for
             full-length clones and constructed by Life Technologies.
             Note: this is a NIH_MGC Library."

FEATURES
source
    Location/Qualifiers
    1..953
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4339187"
    /clone_lib="NIH_MGC_85"
    /tissue_type="lymphoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    High quality sequence stop: 704.

BASE COUNT   214 a   255 c   271 g   213 t
ORIGIN

Query Match   26.5%;   Score 656.6;   DB 169;   Length 953;
Best Local Similarity   97.7%;
Matches 687; Conservative 0; Pred. No. 5e-161; Mismatches 14; Indels 2; Gaps 2;

```
Qy   594 gagtccagaaaggcctctccagcaggctcagtccagagccgatcttcaggagactctc 653
         |||||||||| || |||||||||||||||||||||||| |||| ||||||||||||||
Db     7 GTGGGCAGAAAAGGCCTCAGCCAGCAGCCTCAGTCCAGAGCGATCTTCAGACTCCGAGTTGAA 66

Qy   654 tgaaaatgagccacacctcccacatggttgttagccagagagaagaggggtcaggggactctc 713
         |||||| ||||||||||||||| |||||||||||||||||||||| ||||| ||||||||||
Db    67 TGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAGAGAGGGGTCAGGGGACTCTTC 126

Qy   714 cctggtcgtagcttcatctgtaagctcacttaaagaggaggaaaacttcttggtgctcaa 773
         ||||||||||| |||||||||| |||||||||||| ||||||||| |||||||||| ||||
Db   127 CCTGGTCGTAGCTTCATCTGTAANCTCACTTAAAGAGGAGGAAAACTTCTTGGTGTGCTCAA 186

Qy   774 agcaaaggagatgggccctcccaagttgggacagctgccatctgccctccccatcaattgctgt 833
         |||||||||||||| ||| ||||||||||||||||||||| |||||||||| |||||||||||||
Db   187 AGCAAAGGAGATGGGGCCTCCCAGTTGGGACAGCTGCCATCTGCCCTCCCCATCATTGCTGT 246

Qy   834 caaggacgggaaaagcatcactcatgaaggaagagagagagttttggctgaagagctgtgtac 893
         |||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    ... caaggacgggaaaagcatcactcatgaaggaagagagagagtttttggctgaagagctgtgtac 893
```

```
RESULT 14
BF525432
LOCUS       BF525432      947 bp    mRNA         EST       11-DEC-2000
DEFINITION  602069517F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4212365
            5', mRNA sequence.
ACCESSION   BF525432
VERSION     BF525432.1   GI:11612793
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 947)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9783 row: a    column: 06
            High quality sequence stop: 695.

FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4212365"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.57 kb. Constructed by Life
                     Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT      224 a    257 c    289 g    177 t
ORIGIN
```

```
Query Match    96.6%;  Score 655.8;  DB 150;  Length 947;
Best Local Similarity   96.6%;  Pred. No. 8e-161;
Matches  745;  Conservative  0;  Mismatches  17;  Indels  9;  Gaps  7;

Db   247 CAAGGACGGGAAAAGCATCACTCATGAAGGAAGAGAGAGATTTTGGCTGAAGAGCTGTGAC 306
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   894 tcctccagaatcctggttctgctgtttttgttggtttggttagaatgttccagatgaaagcttcattca 953
         ||||||||||||||||||||||| |||||| ||||||||| ||||||| |||||||||||||
Db   307 TCCTCCAGAATCCTGGTTCTGCTTTTGTTGTGGTTAGAATGTCCAGATGAAAGCTTCATTCA 366
Qy   954 accatctgtgagaatgccaacttttcagaggtaccaaagaggaaagtgaccccgtggc 1013
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 ACCCATCTGTGAGAATGCCAACTTTTCAGAGGTACCAAAGAGGAAAGTGACCCCCGTGGC 426
Qy  1014 cttggtggttcacatggccccagcaatctgtcttgtggacagacgggtaccacgagtggat 1073
         ||||||||||||||||||||||||||||||| |||| ||||||| ||||||||||||||
Db   427 CTTGGTGGTTCACATGGCCCCAGCAATCTGTGTTGCTGTGGACAGCAGGGTACCACGAGTGGAT 486
Qy  1074 ggagaggtttgggcctgacaccaggcaacacttggttcctgaatgagagaactgttcctcaagttca 1133
         ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||| |||||
Db   487 GGAGAGGTTTGGGCCTGACACCAGCAACACTTGGTTCCTGAATGAGAGAACTGTTCCTCAGTTCA 546
Qy  1134 caacccttggcacacaagaattcaaacccagctcaacctcaccccagacacatcttcc 1193
         |||||| || | ||||||||||||||||||||||| |||||||||| ||||||||||||
Db   547 CAACCTTCGCAG-CACACAGAATTCAAAACCCAGCTCAAACCTCACCCGGACACATCTTCCC 605
Qy  1194 cctgtcaccagttccgctgtaagagaaggagggggcccaccccctcagtgtgcccagttca 1253
         |||||||||||| ||||||||||| |||||||| ||||| ||||||| || | |||||||
Db   606 CCTGTCACCAGTTTCCGCTGTAAGAAGAGGAGGGGCCCCACCCCCTCAATGGTCCCAGTTCA 665
Qy  1254 gggtgaatgcctcctcaagtaccagctccgtccaggaggag 1296
         ||||||||| ||||||||||||||||||||||||||||||
Db   666 GGGTGAATG-CTCCTCAAGTACCAGCTCCGTCCAGGAGGAGTG 707
```

```
RESULT 15
BE902696
LOCUS       BE902696      735 bp    mRNA         EST       29-SEP-2000
DEFINITION  601677393F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
            mRNA sequence.
ACCESSION   BE902696
VERSION     BE902696.1   GI:10393148
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
```

```
Query Match   26.4%;  Score 655.8;  DB 150;  Length 947;
Best Local Similarity   96.6%;  Pred. No. 8e-161;
Matches  745;  Conservative  0;  Mismatches  17;  Indels  9;  Gaps  7;

Qy  1380 gcaggagtacagggagtgagtgcgcagagacgggcccagcccagcaggagaaagaagtcagta 1439
         ||||||| || ||| ||||| ||| || ||| ||||||||||||||| ||||| |||||||
Db     1 GCAGGAGTGACAGGAGGAGTGCCGCGAGGACGGCCCAGCCCAGCAGAGAAAAGAAGTCAGTA 60
Qy  1440 cccagaaatcatcttcccttggaacaggttctgccatcccgatgaaagattcgaaatgtcag 1499
         ||||||| ||||||||||||||||||||| ||||||||||||||||||| |||||||||||
Db    61 CCCAGAAAATCATCTTCCCTTGGAACAGGGTTCTGCCATCCCGATGAAAGAATTCGAAATGTCAG 120
Qy  1500 tgccacacttgtcaacataagccccagacacgtctctgtcgtactgtggttggatggttgggac 1559
         ||||||||||||||| |||||||||||||||||||||||||||||| || || | | |||| |
Db   121 TGCCACACTTGTCAAACATAAGCCCCAGACACGTCTCTGTCGTACTGGGACTGTGGTGAGGGCAC 180
Qy  1560 atttgggcagctgtgccgtcattacggagaccaggtggacaggggtcctgggcaccctggc 1619
         | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ANTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGGTCCTGGGCACCCTGGC 240
Qy  1620 tgctgtgttcgtgtccaacctgcacgcagaatcacccaaacggggcttgccaagtatctctgct 1679
         |||||||| | ||||| |||||| ||| ||||||| || ||||| ||||||||||||| |||
Db   241 TGCTGTG-TTGTGTCCCACCTGCAGCGATCACCACAACCCGGGCTTGCCAAGTATCTTGCT 299
Qy  1680 gcagagagaaccgcgccccttggcatctttgggaaacgccgcttcaccccttgctgttggttgc 1739
         |||||||||||||||| |||||||||||||||||| ||||||||||||| |||||||||||||
Db   300 GCAGAGAGAACCGCGCCTTGGCATCTTTGGGAAAGCCCGCTTCACCCTTTGCTGTTGGTTGC 359
Qy  1740 cccaaccaagctcaaagctcggtccagcagtaccacacaacacaagtgccaggaggtcctgca 1799
         |||||||||||||||||| | || || ||||||||||||||| || |||||||||||||| ||
Db   360 CCCAACCAAGCTCAAAGCCTGGCTCCAACCAGTACCACACAACCACCAGTGCCAGGAGGTCCTCA 419
Qy  1800 ccacacaagtatgattccctgccaaatgccttcaggaaggggctgagatctccagatcctgc 1859
         ||||||||||||||| ||||||||||||||||||||||||||||||||||| ||||| ||||
Db   420 CCACACAAGTATGAATTCCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCCTCCAGTCCTGC 479
Qy  1860 agtggaagaatgatcagttccgtgttgcgaacaatgtgaatttggaaggagttcaagactg 1919
         ||||||| |||| |||||||||||||||||||||||||| ||||||| ||||| |||| ||
Db   480 AGTGGAAAGAATTGATCAGTTCCGTGTTGCGAACAATGTGA-TTGGAAGAGTTTCAGACCTG 538
Qy  1920 tctggtgcggcactgcaagcatgcgttggctgtgctggttgcacacctctggctggaa 1979
         ||||||||||||||||||||||||| |||||||||  || ||||||||||||||||||
Db   539 TCTGGTGCGGCACTGCAAGCATGCG-TTGGCTGTGCGCCT-GTGCACACCTCTGGCTGGAA 596
Qy  1980 agtggtctattccgggagacaccatgccgctgcgagggtcctggtgtccggga---tggggaaaga 2036
         ||||||||||||||||| |||||||||||||||||||||||||| ||||||||   || |   |
Db   597 AGTGGTCTATTCCGGGGACACCATGCCGCTGCGAGGGTCCTGGTCGTCCGGAAGA 656
Qy  2037 tgccacctcctgatacatgaagccacaccctgaaagtgggttggaagaggaagcagtgga 2096
         ||||| ||||||||||| ||||||||||| || ||||||| ||||||||||||| |||||
Db   657 TGCCAGCCTCCTGATACAATGAAGCCACACCCCTGGGAAGTGG-TTGGAAGAGGAAGCCAGTGGA 715
Qy  2097 aaagacacacagacacaagtcccaagccatcagcgtggggatgccggatgaa 2147
         ||||| |||| || ||||| ||||||| ||||| |||||| ||| |||||||
Db   716 AAGACCCACCAGCACAA-GTCCCAAGCATCAAGCGGGGGGGATCGGAGATGAA 765
```

```
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
           Plate: LLCM841 row: j column: 23
           High quality sequence stop: 732.
FEATURES            Location/Qualifiers
     source         1..735
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:3959926"
                    /clone_lib="NIH_MGC_21"
                    /tissue_type="choriocarcinoma"
                    /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                    Site_2: EcoRI; cDNA made by oligo-dT priming.
                    Directionally cloned into EcoRI/XhoI sites using the
                    following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                    for average insert size 1.8kb. Library constructed by
                    Ling Hong in the laboratory of Gerald M. Rubin (University
                    of California, Berkeley) using ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      168 a     206 c     203 g     157 t     1 others
ORIGIN

Query Match      26.4%;   Score 655.6;   DB 141;   Length 735;
Best Local Similarity   97.4%;   Pred. No. 8.4e-161;
Matches   719;   Conservative   0;   Mismatches   14;   Indels   5;   Gaps   5;

Qy      1190 tcccctgctcaccagttccgctgtaagaaggaggcccccacctcagtgtgcccatgg 1249
             ||||||||||||||||||||||||||| |-|||||||||||||||||||||||||||
Db         1 TCCCCCTGCTCACCAGTTTCCCGCTGTAAGAAGGAGGCGCCACCTCAGTGTGCCCATGG 60

Qy      1250 ttcagggtgaatgcctccagtaccagctccgtcccaggagaggagttggcagagggatg 1309
             ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TTCAGGATGAATGCCTCCCTCAAGCTACCAGCTCCGTCCCAGGAGAGGAGTTGGCAGAGGGATG 120

Qy      1310 ccattattactt-gcaatcctgaggaattcatcagtgtgaggcgctgcagcttcccaacttc 1368
             ||||||||||   |||||||||||||||||| |||||||||||||||||||||||||||||
Db       121 CCATTATTACTTNGCAATCCTGAGGAATTCATCAGTTGAGGCGCTGCAGCTTCCCAACTTC 180

Qy      1369 cagcagagcgctgcaggagtacagaggaggagtgcgcaggacggcccagccccaagagaa 1428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAAGAGAAA 240

Qy      1429 agaagtacagtaccacagaaatcatcctccttggaacagggtctgcacacccgatgaagatt 1488
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AGAAGTACAGTACCACCAGAAATCATCTTCCTTGGAACACAGGGTCTGCCATCCCGATGAAGATT 300

Qy      1489 cgaaatgtcagtcacacttgtcacacatagcccgacacgtctctgctactggactgt 1548
             ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CGAAATGTCAGTCGCCACACTTGTCACACATAAGCCCGACACGTCTCTGCCTACTGGACTGT 360

Qy      1549 ggtgagggcacatttgggcagctgtccgctcattacggagaccaggtggacagggtcctg 1608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GGTGAGGGCACATTGGGCACGCTGCCGGCTCATTACGGAGACCAGGTGGACAGGGTCCTG 420

Qy      1609 ggcacctgcgctgctgttgtgtccacctgcacgcagcgagtacacacacgggcttgcca 1668
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GGCACCCTGGCTGCTGTG-TTGTGTCCCACCTGCACGCAGCATCACACACGGGCTTGCCA 479

Qy      1669 agtatcttgctgcagagagaacgcgcccttggcaatcttgggaaagccgcttcaccctttg 1728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       480 AGTATCTTGCTCCAGAGAGAGAACGCGCCCTTGGCATC-TTGGGAAAGCCCGCTTCACCCTTTG 538

Qy      1729 ctggtggttgccccaaccagctcaaagctggctcagacagtaccaacacagtgcag 1788
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       539 CTGGTGGTTGCCACACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAAGTC 598

Qy      1789 gaggtcctgcaccacatcagtatgattcctgccaaatgccttcaggaagggctgaagatc 1848
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       599 GAGGTCCTGCCACCACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAAGTC 657

Qy      1849 tccagtcctgcagtggaaagaattgatcagttcgctcgttgccgaaacatgtcgattt-ggaaga 1907
             |||||||||||||||||||||||||| |||||||||||||| ||||||||||||| |||||
Db       658 TCCAGTCCTGCAGTGGAAAGAATTGATCAGTTCGCTCGTTGCCGAAACATGTCGATTTGGGAAGA 717

Qy      1908 gtttcagacctgtctggt 1925
             ||||||||||||||||||
Db       718 GTTTCAGAACTGTCTGGT 735

Search completed: October 28, 2001, 18:35:36
Job time: 6215 sec
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: October 28, 2001, 18:35:41 ; Search time 3261.8 Seconds
                                            (without alignments)
                                            11765.125 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgcttgctcgct..........agaaggtcagagagccccagtga 2481

Scoring table: OLIGO_NUC
               Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size :   8

Total number of hits satisfying chosen parameters: 2427854

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   GenEmbl:*
        1: gb_ba1:*
        2: gb_ba2:*
        3: gb_ba3:*
        4: gb_in1:*
        5: gb_in2:*
        6: gb_in3:*
        7: gb_om:*
        8: gb_ov:*
        9: gb_pat1:*
       10: gb_pat2:*
       11: gb_ph:*
       12: gb_pl1:*
       13: gb_pl2:*
       14: gb_pl3:*
       15: gb_pl4:*
       16: gb_ba1:*
       17: gb_ba2:*
       18: gb_fun:*
       19: em_htgo_hum:*
       20: em_htgo_inv:*
       21: em_htgo_rod:*
       22: em_htg_hum1:*
       23: em_htg_hum2:*
       24: em_htg_hum3:*
       25: em_htg_hum4:*
       26: em_htg_hum5:*
       27: em_htg_hum6:*
       28: em_htg_hum7:*
       29: em_htg_hum8:*
       30: em_htg_inv1:*
       31: em_htg_inv2:*
       32: em_htg_other:*
       33: em_htg_rod:*
       34: em_hum1:*
       35: em_hum2:*
       36: em_hum3:*
       37: em_hum4:*
       38: em_hum5:*
       39: em_hum6:*
       40: em_hum7:*
       41: em_in:*
       42: em_om:*
       43: em_or:*
       44: em_ov:*
       45: em_pat:*
       46: em_ph:*
       47: em_pl:*
       48: em_ro:*
       49: em_sts:*
       50: em_sy:*
       51: em_un:*
       52: em_vi:*
       53: gb_sts1:*
       54: gb_sts2:*
       55: gb_sts3:*
       56: gb_sy:*
       57: gb_un:*
       58: gb_vi1:*
       59: gb_vi2:*
       60: gb_htg1:*
       61: gb_htg2:*
       62: gb_htg3:*
       63: gb_hcg4:*
       64: gb_htg5:*
       65: gb_htg6:*
       66: gb_htg7:*
       67: gb_htg8:*
       68: gb_htg9:*
       69: gb_htg10:*
       70: gb_htg11:*
       71: gb_htg12:*
       72: gb_htg13:*
       73: gb_htg14:*
       74: gb_htg15:*
       75: gb_htg16:*
       76: gb_htg17:*
       77: gb_htg18:*
       78: gb_htg19:*
       79: gb_htg20:*
       80: gb_htg21:*
       81: gb_htg22:*
       82: gb_htg23:*
       83: gb_htg24:*
       84: gb_htg25:*
       85: gb_pr1:*
       86: gb_pr2:*
       87: gb_pr3:*
       88: gb_pr4:*
       89: gb_pr5:*
       90: gb_pr6:*
       91: gb_pr7:*
       92: gb_pr8:*
       93: gb_pr9:*
       94: gb_ro1:*
       95: gb_ro2:*
       96: gb_in4:*
       97: gb_pr10:*
       98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	2481	100.0	2908	89	AF304370	AF304370 Homo sapi
2	2379	95.9	2997	91	BC001939	BC001939 Homo sapi
3	2379	95.9	3006	91	BC004158	BC004158 Homo sapi
4	2022	81.5	2976	89	AK001392	AK001392 Homo sapi
5	1667	67.2	2908	89	AF308698	AF308698 Pan trogl
6	1266	51.0	2893	89	AF308694	AF308694 Gorilla g
7	247	10.0	118788	85	AC005277	AC005277 Homo sapi
8	142	5.7	740	89	AF304371S2	AF304369 Homo sapi

```
RESULT   1
AF304370
LOCUS       AF304370    2908 bp    mRNA            PRI       23-FEB-2001
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 mRNA, complete cds.
ACCESSION   AF304370
VERSION     AF304370.1  GI:10880932
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2908)
AUTHORS     Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE       A candidate prostate cancer susceptibility gene at chromosome 17p
JOURNAL     Nat. Genet. 27 (2), 172-180 (2001)
PUBMED      11175785
REFERENCE   2  (bases 1 to 2908)
AUTHORS     Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
TITLE       Direct Submission
JOURNAL     Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     CDS             1..2481
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24441.1"
                     /db_xref="GI:10880933"
                     /translation="MWALCSLRSAAGRTMSQGRTISQAPARERERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLITRMWSNVGGLSGMTLTLKETGLPRCVLSGPPQLEKYLERAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRGKHQPWQSPRRPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRGRVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVRDGKSITHEGREILAEELCTPPDGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQOWMERFGPDTOHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPREWRQRDAIITCNPEEFIVEA
                     DTSLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKAWLOOYHNOCOEVLHHISMIPAKCLOEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHYSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEAAVEKTHSTTSQAISVGMRNAEFTMLNFHFSQRYAKVPFLFSPNFS
                     EKVGVAEFDHMKVCFGDFPTMPKLIPFLKALFAGDIEEMEBERREKRELRQVRAALLSRE
                     LAGGLEDGEPOQOKRAHTEEPOAKKVRAO"
BASE COUNT      702 a    793 c    826 g    587 t
ORIGIN
```

ALIGNMENTS

```
   9  122       89  AF304371S1  AF304371  Homo sapi
  10   47  1.9 2712  94  AF308696    AF308696  Mus muscu
  11   46  1.9 34593 94  AF348157    AF348157  Mus muscu
c 12   23  1.9 43514 86  AC005545    AC005545  Homo sapi
  13   23  0.9 105490 89 AF165142    AF165142  Homo sapi
  14   23  0.9 190705 65 AC019331    AC019331  Homo sapi
  15   23  0.9 64151  90 AL139399    AL139399  Human DNA
c 16   21  0.9 125403 87 AC008929    AC008929  Homo sapi
c 17   21  0.8 147807 70 AC027347    AC027347  Homo sapi
  18   21  0.8 162344 91 AL442223    AL442223  Homo sapi
  19   21  0.8 196424 91 AP000719    AP000719  Homo sapi
  20   21  0.8 196613 83 AP002490    AP002490  Homo sapi
  21   21  0.8 220146 71 AC034179    AC034179  Homo sapi
  22   20  0.8 1176    8  AB046207    AB046207  Anguilla
  23   20  0.8 1573   14  U23796      U23796    Arabidopsis
  24   20  0.8 4918   12  AF047031    AF047031  Arabidops
  25   20  0.8 6718   85  AB014516    AB014516  Homo sapi
  26   20  0.8 13800   3  SC7B7       AL009199  Streptomy
c 27   20  0.8 46275  85  AC003107    AC003107  Human DNA
  28   20  0.8 79550  12  AB007727    AB007727  Arabidops
  29   20  0.8 84129  12  AB013392    AB013392  Arabidops
  30   20  0.8 103638 13  ATF14F18    ATF14F18  Arabidops
  31   20  0.8 114521 90  AL159985    AL159985  Human DNA
  32   20  0.8 117338 67  HS173D1     HS173D1   Human DNA
c 33   20  0.8 127953 57  AC022835    AC022835  Homo sapi
c 34   20  0.8 148508 71  AC027795    AC027795  Homo sapi
c 35   20  0.8 149006 88  AC025812    AC025812  Homo sapi
c 36   20  0.8 149180 63  AC015525    AC015525  Homo sapi
c 37   20  0.8 150683 74  AC069420    AC069420  Homo sapi
c 38   20  0.8 165490 65  AC025326    AC025326  Homo sapi
c 39   20  0.8 165493 65  AC018876    AC018876  Homo sapi
c 40   20  0.8 165914 81  AL441923    AL441923  Homo sapi
c 41   20  0.8 166344 79  AL162577    AL162577  Homo sapi
c 42   20  0.8 168997 66  AC021927    AC021927  Homo sapi
c 43   20  0.8 177868 69  AC025191    AC025191  Homo sapi
c 44   20  0.8 183412 83  AP001899    AP001899  Homo sapi
c 45   20  0.8 183855 61  AC009591    AC009591  Homo sapi
```

Query Match   100.0%;   Score 2481;   DB 89;   Length 2908;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 2481;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy     1  atgtgggcgcttgctttgctctgctgcggtccgaggtccgaggtccggacgcaccaatgtcgcaggggacgc   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGTGGGCGCTTGCTTTGCTCGTCCGGGTCC...gtccgaggtccggacgcaccaatgtcgcaggggacgc   60

Qy    61  accatatcgcaggcgacccgccccgcgcgccgcagcgcgccgcaagacccggccctgccgggcaacctg  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  ACCATATCGCAGGCACCCGCCCCGCGCGCCGAGCGCGCCGCAAGACCCGGCCCTGCCGGGCAACCTG  120

Qy   121  cgcacgcgagagaagacgcggaccgtcgggtgtctccgggcgggcccaaaacccgtgtacctg  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CGCACGCGAGAGAAGACGCGGACCGTCGGGTGTCTCCGGGCGGGCCCAAAACCCGTGTACCTG  180

Qy   181  cagggtgttggcagcgggttagcgggggactcggggactcggggactcggggactcggggactc...  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CAGGTGGTGGCAGCGGGTTAGCGGGGGACTCGGG...agagagagcgt...tccgagttc  240

Qy   241  aaccggtatcttcttcaacctgtgggagaaggcgtcagagagactcatgtcaggagacacaagtta  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  AACCGGTATCTTCTTCAACCTGTGGAGAAGGCCGTTCAGAGACTCATGCAGGAGCACAAGTTA  300

Qy   301  aaggttgctcgcctggacaacatattcctgacacgaatcgcactggtctaatgttgggggc  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  AAGGTTGCTCGCCTGGACAACATATTCCTGACAGAATCGCACTGGTCTAATGTTGGGGGC  360

Qy   361  ttaagtggaatagattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  TTAAGTGGAATAGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGA  420

Qy   421  cctcccacaactggaaaaaataccctgaagcaatcaaaaatattttctggtccattgaaagga  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  CCTCCACAACTGGAAAAAATACCTCGAAGCAATCAAAAATATTTTCGGTCCATTGAAAGGA  480
```

```
Qy    481  atagaactggctgtgtcgcggcccactctgcccagaaatacgaggatgaaacctgacagtt  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  ATAGAACTGGCTGTGTCGCGGCCCACTCTGCCCAGAAATACGAGGATGAAACCTGACAGTT  540

Qy    541  taccagatcccatacacagtgaacagaggagggaaagcaccaaccatggcagagtcca     600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TACCAGATCCCATACACAGTGAACAGAGGAGGGAAAGCACCAACCATGGCAGAGTCCA     600

Qy    601  gaaaggcctcagaggctcagtccgagctcagactccgagtcgaatgaaat            660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGTCGAGAATCTTCAGCAGAGTCCA         660

Qy    661  gagccacaccttccaacatggtgttagccagagaagagaggggtcaggggactcttccctggtc  720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GAGCCACACCTTCCACACATGGTGTTAGCCAGAGAAGAGAGGGGTCAGGGGACTCTTCCCTGGTC  720

Qy    721  gtagcttcatctgtaagcttcacttaaagagaggagaagcaaagaaag              780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAGAAGCTTCAGAAGCAAAG         780

Qy    781  gagatgggcctcccagtgttgggacagtgccatcgctcccaatcattgctgtcaaggac       840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  GAGATGGGCCTCCCAGTGTTGGGACAGCTGCCCATGGCTCCATCATTGCTGTCAAAGCAAAG     840

Qy    841  gggaaaagcactcactcatgaagaagaagagattttgggctgaagaagctgtgtactcctcca     900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  GGGAAAAGCATCACTCATGAAGAAGAAGAGATTTTGGGCTGAAGAGCTGTGTACTCCTCCA      900

Qy    901  gatcctggttgctgcttttgtgtggttagaaagcttcaacccatc                960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GATCCTGGTTGCTGCTTTTGTGTGGGTTAGAAATGTCCAAGATGAAAAGCTTCAACCCATC    960

Qy    961  tgtgagaatgccaccttttcagaggggttaccaacagatgccccgctgttggtg        1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  TGTGAGAATGCCACCTTTCAGAGGTACCCAAGGAAAAGGCAGTGCCCCGCTTGGTG        1020

Qy   1021  gttcacatggcccacagcatctgtgctgttgtgaaacagtgtggatgggagagg       1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  GTTCACATGGCCCACAGCATCTGTGCTGTTGTGACAGCACAGTGTGGATGGGAGAGG      1080

Qy   1081  tttgggcctgacaacccagcacttggtctccagttcagggtga            1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  TTTGGGCCTGACACCCAGCACTTGGTCTCTGAATGAGAACTGTGCCTCAGTTCACAACCTT    1140

Qy   1141  cgcagcacagattcaaacccagctcaaactcatcagaggcacctctc            1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  CCCGAGCCACACAGATTCAAAACCCAGCTCAACCTGATCCTCATCTCCCCTGCTC        1200

Qy   1201  accagttccgctgtaggaaactcagtgttgcccatggttcagggtgaa             1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  ACCAGTTTCCGCTGTAAGAAGAGGACGCCCACCTCAGTGTGTGCCATGGTTCAGGGTGAA    1260

Qy   1261  tgcctcctcaagtaccagcctcgtcccaggagggagtggcagggatgccatt attact    1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  TGCCTCCTCAAGTACCACCTCCGTCCCAGGAGGGAGTGGCAGGGATGCCATTATTACT     1320

Qy   1321  tgcaatcctgaggaattcatatgttgaggcgctgcagagcttcccaacttccagcagacgtg    1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  TGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGAGCTTCCCAGGAGGGAGGCCA      1380

Qy   1381  caggagtaacagaggagtgcgaggaccagggcccagcccagagagagaaagaaagtcagtac    1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  CAGGAGTACAGAGGAGTGCGAGGACCAGGGCCCAGCCCAGAGAGAGAAAGAAAGTCAGTAC    1440

Qy   1441  ccagaaatcatcttcctggaacaaggtgctgccatccgagtgaaagattcgaaatgtcagt    1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCGAGTGAAAGATTCGAAATGTCAGT    1500

Qy   1501  gccacacttgtcaacataaagccccgacactctcttgctactggactgttggtgaggcaca    1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  GCCACACTTGTCAACATAAAGCCCCGACAGTCTCTGCTACTGGACTGTTGGTGAGGGCACA    1560
```

```
Qy   1561  tttgggcagtgtgccgtcattacggagaccaagttggtggacaggggtcctgtggcaccctggct  1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  TTTTGGGCAGCTGTGCCGTCATTACGGAGACCAAGGTTCCTGGGACCAGGGTCCTGTGGGCCACCTGGCT  1620

Qy   1621  gctgtgtgttgtgtccaactgccacgccagatcaaccaccaggttgccaagtatcttgctg  1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  GCTGTGTTGTTGTGTCCACACTGTGCACGCCAGATCACCACCACGCGGCTTGCCAAGTATCTTGCTG  1680

Qy   1681  cagagagagaaacgcgccttggcatcttcttgggaaagcagtcaccttcatcccttgctggtgttgcc  1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  CAGAGAGAAACGCGCCTTGGCATCTTTGGGAAAGTCACCCTTTGCTGGTGGTTGCC   1740

Qy   1741  cccaaccagtcaaagctcaaagcctggctccagcagtaccacaccagtgccagggaggtctgggcac  1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  CCCAACCAGTCAAAGCCTGGCTCCAGCAGTACCACAACCCTTTGCTGGTGGTGGTTGGCTCAC  1800

Qy   1801  cacatcagtatgattcagttcgctgttcggaacaatgtccttcaggaggaggctgagagatctccagttcctgca  1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  CACATCAGTATGATTCAGTTCGCTGTTCCGAAATGCCTTCAGGAGGAGGCTGAGATCTCCAGTTCCTGCA  1860

Qy   1861  gtggaaagattgatcagttcgctgttgtcgcatgtcgctgtgtcgtggaaacctctgctggaaa  1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  GTGGAAAGATTGATCAGTTCCTGCAAATGCCTTCAGGACTTGGAACATTGATTTGAAGACCTGT  1920

Qy   1921  ctggttgcggcactgaagctgcaagctgttgtcgctgtggtctgtcaacctctgctggaaa  1980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  CTGGTTGCGGCACTGCAAGCTGCAAGCGTTTGGCGTGCTGGTCGTGTCACACCCTCTGGCTGGAA  1980

Qy   1981  gtggtctattccgggacacatgcccgggaggctcggtccggacttggggaaaagtgcc  2040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  GTGGTCTATTCCGGGACACATGCCCGGGAGGCTCGGGTCCGGACTTGGGGAAAGATGCC  2040

Qy   2041  accctcctgatacaatgaagtgggaaatgtgttcggaaagtgaaaagaagagagaaagaaggtgaaaag  2100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  ACCCTCCTGATACATGAAGTGGGAAATGTGTTTGGGAAAGTGAAAGAAGAGGAAAGGAGAAGCAGTGGAAAAG  2100

Qy   2101  acacacagcacaacgtcccaagccatcagcgtgtggggatgcggaacgcggagttcatt  2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  ACACACAGCACAACGTCCCAAGCCATCAGCGGTGTGGGGATGCGGAACGCGGAGTTCATT  2160

Qy   2161  atgctgaaacacttcaagcagcgctatgccaaggctccaccctcatctcaagc  2220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161  ATGCTGAAACACTTCAAGCAGCGCTATGCCAAGGCTCCACCCTCATCTCAAGC  2220

Qy   2221  gagaaagtgggagttgcctttgacttttgggacacttttccaacaatg          2280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  GAGAAAGTGGGAGTTGCCTTTGACTTTGGGACACTTTTCCAACAATG        2280

Qy   2281  cccaagctgattccccactgaaagcccctgtttgtgcggacatggaggagagaatgaggag  2340
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  CCCAAGCTGATTCCCCACTGAAAGCCCTGTTTGTGCGGACATCGAGGAGAGAATGAGGAG  2340

Qy   2341  ggcaggagagaagcgggagctgcgggagtcgcggggtggcggcagtgcgggg  2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  CGCAGGGAGAGAAGCGGGAGCTGCGGGAGTCGCGGGGTGGCGGCAGTGCGGGG  2400

Qy   2401  ggcggggctgaggggatcggggaggcctcagcgaggaagggagccacaggcc  2460
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401  GGCGGGGCCTGGAGGATGGGGAGGGCCCAGCGAGGAAGGGAGACCCACAGGCC  2460

Qy   2461  aagaaggtcagagcccagtga  2481
           |||||||||||||||||||||
Db   2461  AAGAAGGTCAGAGCCCAGTGA  2481
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2997)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca

            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: 1 Column: 17.
FEATURES             Location/Qualifiers
     source          1..2997
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:4102 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             61..2541
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH01939.1"
                     /db_xref="GI:12804973"
                     /translation="MWALCSLLRSAAGRTMSQGRRISQAPARRERPRKDPLRHLRTRE
                     KRGPSGGCSGPNTVLQVVAAGSRDSGAALYVFSEFNRLFNCGEGVORLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILITLKETGLPKCVLSGPFQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEORGKHQPWQSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRCGVRDSSLVVARICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDGAAFVVVECDESFIQPICENATFORYQGRAD
                     APVALVVHNAPASVLVDSRYQOVMERFGPDTQHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIFPPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSILLDCCEGTFGQLCRHYGDQDVRVLGTLAAYFVSHLHADHHTGLPSILIQRERAL
                     ASLGKPLHPLLVVARPNQLKAWLOOYHNOCOEVLHIHISMIPAKCLOEGAEISSPAVERL
                     ISSLLRTCDLEEFQCLVRHCKHAFGCALVTHTSGWKVVYSGDTMPCEALVPMGKDATYL
                     LIHEATLEEDGLEEAVEKHSHTTSOAISVGMRMNAEFIMLNHFSQRYAKYPLFSPNFS
                     EKVGVAFDHMKVCFCDFPPTMPKLIPPLKALFAGDIEEMERRERKELRQVRAALLSRE
                     LAGGLEDGEPQOKRAHTEEPQAKKVRAQ"
BASE COUNT      740 a    806 c    853 g    598 t
ORIGIN
```

```
Query Match             95.9%;  Score 2379;  DB 91;  Length 2997;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2479;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

Qy    1 atgtgggcgcttgctgggtccgggcacccgcgaccacccatgtcgcagggacgc 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATGTGGGCGCTTTGCTCCGTTGCTGCGGGTCCGCGGCCCGGACGACCATGTCGCAGGGACGC 120

Qy   61 accatatcgcaggcgaccccgcgcgggtcgggcgggcggcaaggaccccgctgcggcacctg 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACCATATCGCAGGCGACCCCGCGCGGGTCGGGCGGGCGGCAAGGACCCCGCTGCGGCACCTG 180

Qy  121 cgcacgcgagagaaagcgcggagcacgtcgggtgctcctcggggcgggcccaaacacgctg 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CGCACGCGAGAGAAAGCGCGGAGCACGTCGGGTGCTCCTCGGGGCGGGCCCAAACACCGCTG 240

Qy  181 caggtggttggcagcgggtagcgcgggcgggccgccggctcacgtcttctccgagttc 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CAGGTGGTTGGCACGCGGGTTAGCGCGGGACTCGGGCGGCCGCCGGCTCACGTCTTCTCCGAGTTC 300

Qy  241 aaccggtatctcttcaactgtggagaagcgttcagagactcatgcagcgacacaagtta 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 AACCGGTATCTCTTCAACTGTGGAGAAGCGTTCAGAGACTCATGCAGCGACACAAGTTA 360

Qy  301 aaggttgctcgcctggacacaacatattcctgacacgaatgactggttctaattgttggggc 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGGTTGCTCGCCTGGACACAACATATTCCTGACACGAATGACTGGTTCTAATTGTTGGGGGC 420

Qy  361 ttaagtggaatgattcttacttttaaagaaaccgggcttccaaagtgtgtactttctgga 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 TTAAGTGGAATGATTCTTACTTTAAAGGAAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 480

Qy  421 cctccacaactggaaaaaatacctcgaagcaatcaaaatatttcctggtccattgaaaga 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CCTCCACAACTGGAAAAAATACCTCGAAGCAATCAAAAATATTTCTGGTCCATTGAAAGGA 540

Qy  481 gagccacaaccccaactcctggttagccaagaagagggttcagggactcttccctggtc 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GAGCCACAACCTCCACATGGTGTTAGCCCAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 600

Qy  541 taccagatcccccatacacagtgaacacagaggaggggaaagtcacaaaccatggcagagtcca 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TACCAGATCCCCCATACACAGTGAACACAGAGGAGGGGAAAGTCACAAACCATGGCAGAGTCCA 660

Qy  601 gaaaggcctctcagcaggctcagtccagatcttcagaactccagtcgaatgaaaat 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GAAAGGCCTCTCAGCAGGCTCAGTCCATCGAGTCCAGTCGGATCTTCAGACTCCGAGTTGAATGAAAAT 720

Qy  661 gagccacaactccaactggttcttagccaaggacaaagaagggttcagggactcttcccctggtc 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GACGCCACAACTTCCACATGGTGTTAGCCCAGAAGAGAGGGTCAGGGACTCTTCCCTGGTC 780

Qy  721 gtagcttcatctgtaagcttcacttaaagagaaaacttcttgttgctcaaagcaaag 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GTAGCTTTCATCTGTAAGCTTCACTTAAAAGAGAAAACTTCTTGGTTGCTCAAAGCAAAG 840

Qy  781 gagatgggcgcctccccagttgggacagctgccatcgctcctccccatccattgctgtgtcaaggac 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 GAGATGGGGCGCCTCCCCAGTTGGGACAGCTGCCATCGCTCCTCCCCATCATTGCTGTGTCAAGGAC 900

Qy  841 gggaaaagcatcactcatgaaggagagaagatttgctgagaagctgtgttactcccca 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 GGGAAAAGCATCACTCATGAAGGAGAGAAGATTTTGGCTGAGGAGCTGTGTACTCCTCCA 960

Qy  901 gatcctgtgtgcttttgtggtagaaatgtcccagatgaaagctcattcaacccatc 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 GATCCTGTGTGCTGCTTTTGTGGTGGTTGAAAGTCCAGATGAAAGCTCATTCAAACCCATC 1020

Qy  961 tgtgagaatgccaactttcagaggtaccaaagaagcagtgccccgtggccttggtg 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 TGTGAGAATGCCAACTTTCAGAGGTACCAAGGAAGCAGATGCCCCGTGGCCTTGGTG 1080

Qy 1021 gttcacatggccccagcaatctgtcttgtgttggacagcaggttaccagcagtgggatgggagg 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GTTCACATGGCCCCAGCAATCTGTGTGCTTGTGTTGGACAGCAGGTACCAGCAGTGGGATGGGAGG 1140

Qy 1081 tttgggcctgacaccagcacttggtcctgaatgagaactgccctcagttacaacctt 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1200
```

```
Qy   1141 cgcagccacaagagattcaaaacccagctcaatcttccccctgctc              1200
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGCAGCCACAAGAGATTCAAAACCCAGCTCAATCTTCCCCCTGCTC              1260

Qy   1201 accagttccgctctgttaagaaggaggggcccaccctcagtgttcagggtgaa       1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ACCAGTTCCGCTGTAAGAAGGAGGGGCCCACCCTCAGTGTTCAGGGTGAA          1320

Qy   1261 tgcctcctcagtacagcctccgtccagagagaggaggcggatgccattattact     1320
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGCCTCCTCCAAGTACCAGCTCCGTCCAGAGGAGGAGGCCATTATTACT          1380

Qy   1321 tgcaatcctgaggaattcatagttgtgaggctgcagctccccaacttccagaaggtg    1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGCAATCCTGAGGAATTCATAGTTGTGAGGCTGCAGGGGCCATTATTACT          1440

Qy   1381 caggagtacaggaggagtgcgcagaggaccaggcccccagaggagaaaaagaagtcagtac 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CAGGAGTACAGGAGGAGTGCGCAGAGGACCAGGCCCCCAGAGGAGAAAAAGAAGTCAGTAC 1500

Qy   1441 ccagaaatcatcttccttggaacaagggtctgccatccgatgaaagattcaagtgtcagt  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCAGAAATCATCTTCCTTGGAACAAGGGTCTGCCATCCGATGAAGATTCGAAATGTCAGT   1560

Qy   1501 gccacacttgtcaacataagccccgaacgtctctgtactgctggtgaggggcaca       1560
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCCACACTTGTCAACATAAGCCCCGACGTCTCTGTACTGCTGGTGAGGGGCACG       1620

Qy   1561 tttgggcagctgtgcgtcattacggagggacaagggtctggggacacctggct        1620
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TTTGGGCAGCTGTGCCGTCATTACGGAGGGACAAGGGTCCTGGGCACCCTGGCT       1680

Qy   1621 gctgtgtttgtgtccacctgacgcagagatcaccacacgggcttgccaagtatcttgctg  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GCTGTGTTTGTGTCCCACCTGACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCT    1740

Qy   1681 cagagagaacgcgcctggcatctcttgggaaagcgttcaccccttggctggttgcc      1740
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CAGAGAGAAACGCGGCCTTGGGCCAATCTTTGGGAAAGCGTTCACCCTTGCTGGTTGCC    1800

Qy   1741 cccaacaagctcaaagcctggctgctaccacaacagtgccaggaggttcctgcac       1800
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CCCAACAAGCTCAAAAGCCTGGCTGCTACCACAACCAGTGCCAGGAGGTTCCTGCAC    1860

Qy   1801 cacatacagtagatcctgccaaatgccttcaggaaggggctgagatctccagtcctgca   1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CACATCAGTAGATCATTTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTCCA 1920

Qy   1861 gtggaaagaattgatcagttcgctgttgcgaacatgtgatttggaaagatttcaagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT  1980

Qy   1921 ctggtgggcactgcaagcatggcgtttggcctgtcgcctggtgcaacactctggctggaa  1980
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 CTGGTGCGGCACTGCAACCATGGCGTTTGGCCTGTCGCCTGGTGCAACACTCTGGCTGGAAA 2040

Qy   1981 gtggtctcattccgggacaacatgccctgcgaggctgctgttccggatctgaggaaagtgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GTGGTCTCATTCCGGGACACCATGCCCTGCGAGGCTGCTGTTCCGGATCTGGTGGGAAAGTGCC 2100

Qy   2041 accctcctgataccatgaagcacacctggaagatggtgttggaggaggaagcagtggaaag  2100
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACCCTCGTGATACATGAAGCCACACCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAG   2160

Qy   2101 acacacagcacaacgtcccaagccataagcgtggggatggaatgaacgcggagttcatt   2160
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGGAATGAACGCGGAGTTCATT    2220

Qy   2161 atgctgaaccactcagccacgctatgccaggtcccccctcttcagcccaacttcagc     2220
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATGCTGAACCACTTCAGCCACGCTATCGCCAGGTCCCCCCTCTTCAGCCCCAACTTCAGC   2280

Qy   2221 gagaaagtgggagttgcctttgaccacaatgaaggtctgtgtttgagagacttcccaacaatg  2280
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GAGAAAGTGGGAGTTGCCTTTGACCACACATGAAGGTCTGCTTGCTTGGGACATGGAGGAGG   2340

Qy   2281 cccaagctgattccccactgaaagcccctgtttgctggtggacatcgaggagatggaggag  2340
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CCCAAGCTGAATTCCCCCACTGAAAGCCCCTGTTTGCTGGTGGACATCGAGGAGATGGAGGAG 2400

Qy   2341 cgcaggggagaagcgggggagctgcggggaggctgcggcgggggccctctccgtgggaggtgtggcca 2400
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CGCAGGGGAGAAGCCGGGAGCTGCGGGGAGGCTGCGGCGGGGGCCCCTCTCCGTCCAGGAGGTGGCA  2460

Qy   2401 gggcggcctggaggatgggggagcctcagccagaaagcgggggccaccacagaggagaagcccacagggcc  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GGCGGCCTGGAGGATGGGGGAGCCTCAGCCAGAAAGCGGGGCCACCACACAGAGGAGGACCACAGGCC   2520

Qy   2461 aagaaggtcaagagcccagtga                                       2481
          ||||||||||||||||||||||
Db   2521 AAGAAGGTCAAGAGCCCAGTGA                                       2541
```

RESULT   3
BC004158

LOCUS   BC004158   3006 bp   mRNA   PRI   16-MAR-2001
DEFINITION Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:2441, mRNA, complete cds.
ACCESSION   BC004158
VERSION   BC004158.1 GI:13278770
KEYWORDS   MGC.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3006)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCID/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: c Column: 13.
FEATURES   Location/Qualifiers
     source   1..3006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:2441 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS   70..2550
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH04158.1"
                     /db_xref="GI:13278771"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEEHKLKV

```
BASE COUNT     742 a    808 c    858 g    598 t
ORIGIN

Query Match
Best Local Similarity  95.98%;                           Score 2379;   DB 91;   Length 3006;
Matches 2479;  Conservative  0;   Pred. No. 0;  Mismatches  2;   Indels  0;   Gaps  0;
               99.98%;
```

ARLDNIFLTRMHWHSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELN
ENEPHLPHGYSQSRRGVRDSSLVVAFICKLHLKRGNFLVLRAKEMGLPVGTAAIAPIIA
AVKDGKSITHEGREILAEELCTPPDGCAAFVVVEECPDESFIQPICENATFQRYQGRAD
APVALVWHMAPASVLVDSRYQQWMERGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
HPDIPLPLLTSPRCKKEGPTLSVPMVQGELLKYQLRPRREWRDAIITICNPEEFIVEA
LQLPNFQQSVQEYRRSAQGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
DTSLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFYSHLHADHHTGLPSILLQRERAL
ASLGKPHLPLLVVANQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
ISSLRTCDLEEFQTCLVRHCKHAFCALVHTSGWKVVYSGDTMPCEALVMGKDATL
EKVGVAFDHHKVCFGDFPTMPKLIIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
LAGGLEDGEPQOKRANTEEPAQAKKVRAQ"

Qy   1 atgtgggcgcttcgtcggtcggtccggtccggccaggccaggccaggccagc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 ATGTGGGCGCTTCGTCGGTCGGTCCGGTCCGGCCAGGCCAGGCCAGGCCAGC 129

Qy   61 accatatcgcacggcaccgcccgcccgcgccggcagccgtcgggcaacctg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   130 ACCATATCGCACGGCACCGCCCGCCCGCGCCGGCAGCCGTCGGGCAACCTG 189

Qy   121 cgcacgccgagagaagcgcgaacgtcgggtgtgctccggccaaacacgtacctg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   190 CGCACGCCGAGAGAAGCGCGGACGTCGGGTGTGCTCCGGCCAAACACGTACCTG 249

Qy   181 caggtggtggtcagcggggtagccgggactcgggccgcgcgccctacgctct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   250 CAGGTGGTGGTCAGCGGGGTAGCCGGGACTCGGGCCGCGCGCCCTACGTTC 309

Qy   241 aaccggtatctctcaactgtggagaaggccgttcagagacctcaatgcagacaagtta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   310 AACCGGTATCTCTCAACTGTGGAGAAGGCCGTTCAGAGACCTCAATGCAGACAAGTTA 369

Qy   301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggttctataatgttggggc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   370 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTTCTATAATGTTGGGGC 429

Qy   361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtacttctctgga 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   430 TTAAGTGGAATGATTCTTACTTTAAAGGAAAACCGGGCTTCCAAAGTGTACTTCTCTGGA 489

Qy   421 cctccacaactggaaaaatacctcgaagcaatcaaaaatacctcggagaagcacca 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   490 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAAATATTTCTGGTCCATTGAAAGGA 549

Qy   481 atagaaactgctgtcgcggccccacctcctgcccagaatacgaggatgaaaccatgacagtt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   550 ATAGAACTGCTGTCTGCGGCCCCACCTCGTGCCCAGAATACGAGGATGAAACCATGACAGTT 609

Qy   541 taccagatcccaaccacacagtgaacagagaagcaccaacaccatgcagagtcca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   610 TACCAGATCCCCATACACAGTGAACAGAGAAGCACCAACACCATGCAGAGTCCA 669

Qy   601 gaaaggcctcaagaggctcagtggcaatctcagcacgggtgaatgaaaat 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   670 GAAAGGCCTCTCAGCAGGGCGGATCTTCAGAGCTCAGTGAATGAAAAT 729

Qy   661 gagcacacccttccacatggtgttagtccagagaaggggtcaggggactctttgtc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   730 GAGCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGAGCTCGGTC 789

Qy   721 gtagcttcatctgtaagcttcacttaaagagagaaagaaacttcttggtctgtctcaaaagaaag 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   790 GTAGCTTCATCTGTAAGCTTCACTTAAAGAGAGAAAACTTCTTGGTGTCTCAAAGCAAAG 849

Qy   781 gagatggggcctccagttgggacagctgcctcagctctccaatcattcattcaaggac 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   850 GAGATGGGGCCTCCAGTTGGGACAGCTGCCTCAGCTCTCCATCATTCATTCGTCAAGGAC 909

Qy   841 gggaaaggcatcactcatgaagaaggaagaggaattttgggctgaagggagctgcctcca 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   910 GGGAAAGGCATCACTCATGAAGAAGGAAGAGGAATTTTGGGCTGAAGGGAGCTGTACTCCA 969

Qy   901 gatcctggtgctgcttgtcgttggtggtagaaatgtccaagtttgaaagcttcattcaaccatc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   970 GATCCTGGTGCTGCTTGTTGTTGGTGGTAGAAATGTCCAGATGGAAAAGCTTCATTCAACCCAC 1029

Qy   961 tgtgagaatgccacctttcaagagtgtaccaaagaaaagcagatgcccccgtggccttggtg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1030 TGTGAGGAATGCCACCTTTCAAGAGTACCAAAGGAAAGGCAGATGCCCCCGTGGGCCTTGGTG 1089

Qy 1021 gttcacatggccccacagtctgcttcttgttggacagcaagcagtaccagctggatggaagag 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1090 GTTCACATGGCCCCACAGTCTGCTTCTTGTTGGACAGCAAGCAGTACCAGCTGGATGGAGAG 1149

Qy 1081 tttgggcctgacaccccagcacttggtcctgaattgaaataacagaacttcacaaacctt 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1150 TTTGGGCCTGACACCCCAGCACTTGGTCCTGAATTGAACTGTCGCTTCACAACCTT 1209

Qy 1141 cgcagccacaaagattcaaacccagcacttcaatccatcccagcacaactcatcccagctctc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1210 CGCAGCCACAAAGATTCAAACCCAGCACTCAATCCATCCCAGCACAACTCATCCCAGCTCTC 1269

Qy 1201 accagtttccgctgtaagaaaggagaggggcccacacctcagtgttgccaaccatggtccagggtaa 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1270 ACCAGTTTCCGCTGTAAGAAAGGAGAGGGGCCCACCCTCAGTGTTGCCATGGTTCAGGGTGAA 1329

Qy 1261 tgcctcctcaagtaccagctccgtcccagagaggaggggattggcagagggatggcccattattact 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1330 TGCCTCCTCAAGTACCAGCTCCGTCCCAGAAGGAGGGGGCAGAGGGATGGCCCATTATTACT 1389

Qy 1321 tgcaatcctgaggaattcatagttgtagagcgctcagcttcccaacttccaagagccgtg 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1390 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTCAGCTTCCCAACTTCCAAGAGCCGTG 1449

Qy 1381 caggagtacagggagtgcggcaggacgccccagcgttccagcttcccacagaatcagt 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1450 CAGGAGTACAGGAGTGCGGCAGGACGCCCCAGCTTTCCAGCTACCGGAAGTCAGTAC 1509

Qy 1441 ccagaaatcatcttcttcctggaacagggtctgccatccgatgaagaattcgaaaagtcagt 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1510 CCAGAAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGAATCGAAAGTCAGTAC 1569

Qy 1501 gcaacacttgtcaacataagcccgaacagtctctgtgctactgggttcactggcaca 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1570 GCCACACTTGTCAACATAAGCCCCGACACAGTCTCTGCTACTGGGTTGGTGGGCACG 1629

Qy 1561 tttgggcagctgtgccgtcatacggagacaaggtggacagtgtcacccggct 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1630 TTTGGGCAGCTGTGCCGTCATACGGAGACCAGGTCCTGGGAGGGCACG 1689

Qy 1621 gctgtgttgttgtccaagtgcaatctttgggaaagtctttgcaagtatcttgctg 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1690 GCTGTGTTTGTGTCCAAGTGTGCCGTGCAATCACCACAGGGTTGCCAAGTACTTTGCTG 1749

Qy 1681 cagagagaacgagcccctggcatctctttgggaaagccttcaaccctttgttgttgcc 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1750 CAGAGAGAACGAGCGCCCTGGCATCTTTTGGGAAAGCCTTCAACCCTTTGGTGGTTGCC 1809

Qy 1741 cccaaccagctcaaaggcctggcctggcacagcaccacaacacagggctccgac 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1810 CCCAAACAGCTCAAAAGCCTGGCCTGGCACAGCACCACAACCAGTACCACAAGTGCTCCGAC 1869

Qy 1801 cacatcagtatgatcctgccaaaatgcctcagggaaggggctgaagctctccagtcctgca 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1870 CACATCACTATGATTCCTGCCAAATGCCTTCAGGAGGGGGCTGAGATCTCCAGAGGCTCCGCA 1929

Qy 1861 gtggaagaattgatcagttcgctgttcgcgaacatgtgattggaaagagtttcagacctgt 1920
```

```
Db  1930  GTGGAAAGATTGATCAGTTCGCTGTTGCCAACATGTAGTTGATTGGAAGAGTTCAGACCTGT  1989
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1921  ctggtgtgccgcactgcaacatgcgtgttcggctgtgtgccgctggtgcacactctgtctggaa  1980

Qy  1981  gtggtctattccgggaccaccatgccctcgcgaggctctgtgtccggatggggaaagatgcc  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1990  CTGTGCGCACTCCAAGCATGCCTTTGCTGTCGTCCGCTGGTGCACACTCTGGCTGGAAA  2049

Db  2050  GTGGTGTCATTCCGGGACACCACCATGCCCTCGCGAGGCTCTGTGTCCGGATGGGGAAAGATGCC  2109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2041  accctcctgatacatgaagccacacctggaagatggttgtggaacctggagaaag  2100

Qy  2101  acacacagacaacgtcccaagccaatcagcgtcgtgggatgaagcggggagttcatt  2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2110  ACCCTCCTGATACATGAAGCCACACCTGGAAGATGGTTGTGGAACCTGGAGAAAG  2169

Db  2170  ACACACAGACAACAGTCCCAAGCCAATCAGCGTCGTGGGATGAAGCGGGGAGTTCATT  2229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2161  atgctgaaccacacttcagccaagccgtatgccaagtgtccccctcttcagccccaacttcagc  2220

Qy  2221  gagaaagtggagtgtgcctttgaccacatgaaagtctgctttgtgagagacttctgccaacaatg  2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2230  ATGCTGAACCACTTCAGCCAAGCCGTATCCCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGC  2289

Db  2290  GAGAAAGTGGGAGTTGCCTTTGACCACATGAAAGTCTGCTTTGTGAGAGACTTTCCAACAATG  2349
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2281  cccaagctgattcccccacctgaaagccctgttgctgtgcgactcgagggagatggaggag  2340

Qy  2341  cgcgaggagagagacggggagctcgcggcaggtcgcgggtcccctccgtgtccaaggtggca  2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2350  CCCAAGCTGATTCCCCACTGAAAGCCCTGTTGCTGTGCGACATCGAGGAGATGGAGGAG  2409

Db  2410  CGCAGGGAGAGAGACGGGGAGCTCGCGGCAGGTCGCGGGTCCCTCCGTGTCCAGGAGCTGGCA  2469
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2401  ggcgcgctggggatgggggagcctcagcagaagaaggcggggccacacagaagcacagggcc  2460

Qy  2461  aagaaaggtcagagcgcccaagtga  2481
          |||||||||||||||||||||||||
Db  2470  GGCGCGCTTGGAGGATGGGGGAGCCTCAGCAGAAGAAGGCGGCCACACAGAAGCACAGGGCC  2529

Db  2530  AAGAAAGGTCAGAGCCCAGTGA  2550
```

```
RESULT  4
LOCUS       AK001392     2976 bp    mRNA         PRI     22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION   AK001392
VERSION     AK001392.1  GI:7022621
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2000985.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Kawai,Y., Kimura,J., Wakamatsu,A.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished (2000)
REFERENCE   2  (bases 1 to 2976)
AUTHORS     Isogai,T. and Otsuki,T.
TITLE       Direct Submission
JOURNAL     Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0012, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RP2000985"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pME18SFL3; mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             75..2639
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91666.1"
                     /db_xref="GI:7022622"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERRKDPLRHLRTRE
                     KRGPSGGCSSGPNTVVLQVVAAGSRDSGAALYVFSENRLFNCEGQVQRLMQEHRLKV
                     ARLDNIFLTRMHHWSNVGGLSGMILLKETGLPKCVLSGPPQLEKVLEAIKIFSGPLKG
                     IELAMPHSAPEYEDTMTVYQIPIHSEQRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGYSQRGRVRDSSLVVAFICKIHLKRGNFLVLKAKEMGLPVGTPAAIAPITA
                     AVKDGKSITHEGREIIAEELCTPPDGAAFVVEECPDESFIQPICENATRQRYQGKAD
                     APVALVHMAPASVLVDSRYQQWMERFGDPDQHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIFPLLTSFRCKEGPTLSVPMVGGECLLKYQLRPRENWORDAIITCNPEEFIVEAL
                     QLPNFCQOSVQEYRRSAODGPAPAEERSOYPEIIFLGTGSAIPMKIRNVSATLVNISPD
                     TSLLLDCGEGTFGQLCRHYGDVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERALA
                     SLGKPLHPLLVVAPNOLKAWLQQYVNQCQEVLHHISMIPAKCLGEGAEISSPAVERLI
                     IHEATLEDGLEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVFLFSPNFSE
                     KVGVAFDHMKVCLGDFPTMPKLLIPELKALFAGDIEEMEEREKRELRQVRAALLSREL
                     AGGLEDGSLSRSGPTQRSHRPRSEPSEDLGDPELRRLCVPCPTHAPVSALLAGRS"
BASE COUNT      712 a     807 c     856 g     601 t
ORIGIN
```

```
Query Match             81.5%;  Score 2022;  DB 89;  Length 2976;
Best Local Similarity   99.8%;
Matches 2412;  Conservative 0;  Mismatches 0;  Indels 3;  Gaps 1;

Qy    1  atgtgggcgcacttgctcgctcggtcgggtccgcgggaccgcaacatgtgcaaggggacgc   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75  ATGTGGGCGCACTTTGCTCGCTGCGGTCGCGGGTCCCGCGGGACCGCAACCATGTCGCAGGGGACGC  134

Qy   61  accatacgcaggcaaccgcccgccggaggaggggaaccgctcgggcaactg  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  135  ACCATATCGCAGGCACCGCCCGCCGCGGAGGAGCGCCGGAAGGACCCGCTCGGGCACCTG  194

Qy  121  cgcacacgcgagaagaaagcgggaccgtcggggtgctcgtcgggccaaacacgtgtacctg  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  195  CGCACCCGCGAGAAGAAAGCGGGACCGTCGGGGTGCTCCGCGCCGGCCAAACACGTGTACCTG  254

Qy  181  caggtcggtcggtggcaagcgggtagccgtggacctcgggcgccgtcctacgctccgagttc  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  255  CAGGTCGGTCGGTGGCAAGCGGGTAGCCGTGGACTCGGGGCGCCGTCCTACGCTCTTCCGAGTTC  314

Qy  241  aaccggtatctcttcaacctgtgagaagaagtcaggagacttcaggaggagcaggtta  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  315  AACCGGTATCTCTTCAACTGTGGAGAAGACCGTTCAGGAGACTCATGGAGGAGGAGCAAGTTA  374

Qy  301  aaggttgctcgcctggacaacatattcctgacacacgaatgcactgtgtctaatgttggggc  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  375  AAGGTTGCCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTTGTCTAAATGTTGGGGC  434

Qy  361  ttaagtggaatgattctaactttaaagagaaaccgggcttccaaagtgtgtactttctgga  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  435  TTAAGTGGAATGATTCTTACTTTAAGGAAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA  494
```

```
Qy    421 cctccacaactgtggaaaataccctcgaaagcaatcaaaaatatttctctgtccattgaaagga   480
          |||||||||||||  |||||          |||||||||    || |  |||||   |   ||||
Db    495 CCTTCCACAACTGTTCAACATAAGCCCCGACAAAAATACCTCGAAGCAATCAAAAATATTTTCCATTGAAAGGA   554

Qy    481 atagaactggctgtgtcggcccccactctgccccccagaatacagaaaccatgacagtt   540
          ||||||||| |||||||| ||| ||||||||||||||||||||||||||||||||||||
Db    555 ATAGAACTGGCTATGCGGCCCCCACTCTGCCCCCAGAATACAGAAACCATGACAGTT   614

Qy    541 taccagatccccataccacagtgaacaacaacatggcagagtcca   600
          |||||||||||||||||||||||||||||||||||||||||||||
Db    615 TACCAGATCCCCATACCACAGTGAACAACAACATGGCAGAGTCCA   674

Qy    601 gaaaggccctccagcaagcgctcagtcagtccagacagcagaggggaaggggaaggccca   660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    675 GAAAGGCCCTCTCAGCAGGCTCAGTCCAGACAGCAGAGGGGAAAGCACCAATGGCAGAGTCCA   734

Qy    661 gagccaacctttccacacatggtgttagccaagagaggggagggtcaggggactcttccctggtc   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    735 GAGCCAACCTTTCCACACATGGTGTTAGCCAGAGAGGGGAGGGTCAGGGGACTCTTCCCTGGTC   794

Qy    721 gtagcttcactcgtaagcttcacttaaagagaggaaaaacttcttggttggtcaaagcaaag   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    795 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAAACTTCTTGGTTGGTCAAAGCAAAG   854

Qy    781 gagatggggcctccagcagttggacaagcagctcggccatcagccatcattgctctgtctcaaggac   840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    855 GAGATGGGGCCTCCAGCAGTTGGGACAGCAGTGCCATCAGCCATCATTGTGCTGTGTCAAGGAC   914

Qy    841 gggaaaagcatcactcatgaaggagaagagaagatttggctagcacagaagctcctccca   900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    915 GGGAAAAAGCATCACTCAATGAAGGAAGAGAAGAGTTTGGCTAGCACAGAAGCTCCTCCA   974

Qy    901 gatccctggtgctgtcttcttttgtggtgttagaaatctccaggcaagcttcaaccatc   960
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    975 GATCCTGGTGCTGTCTTCTTTTGTGGTGGGTGTTAGAAATGTCCAGGCAAGCTTCAACCATC   1034

Qy    961 tgtgagaatgccaccctttccagttcacagagtacccagtagtgccccgtggccttggtg   1020
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1035 TGTGAGAATGCCACCCTTTCCAGTTCACAGAGTACCCAGTAGTGCCCCGTGGCCTTGGTG   1094

Qy   1021 gttcacatggccccagcatctgtctgttctgttgtgacagcaagtaccagtggatgaggag   1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1095 GTTCACATGGCCCCAGCATCTGTGCTTGTGTGGCACAGCAAGTACCAGTGGATGAGGAG   1154

Qy   1081 tttgggcctgaacccaacttggtcctgaatgtgcctcagttcacaacctt   1140
          |||||||||||||||||||||||||||||||||||||||||||
Db   1155 TTTTGGGCCTGACACCCAACTTGGTCCTGAATGAGAACTGTCCCTCAGTTCACAACCATC   1214

Qy   1141 cgcagcacaagaattcaaacctcaaacctccaccgacacttccccctgtc   1200
          ||||||||||||||||||||||||||||||||||||||||
Db   1215 CGCAGCACAAGATTCAAAACCCAGCCACCTCAACCTCATTCCACCCGGCACATCTTCCCCGCTC   1274

Qy   1201 accaagttccgctgtaagaggagggccccaccctcagtgtgccccatgttccaggtgttcaaggtga   1260
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1275 ACCAGTTTCCGCTGT---AAGGAGGGCCCCACCCTCAGTGTTCAGGGTGAA   1331

Qy   1261 tgcctcctcaagtacacagctccgtcccacccagaggggaggggatgtcagccattattact   1320
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 TGCCTCCTCAAGTACCACAGCTCCGTCCCACGGTTCCCAGGGATGTCCATTATTACT   1391

Qy   1321 tgcaatcctgaagttggagcggcgctgcagcttccaacttccagcagagcgtg   1380
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 TGCAATCCTGAGGAATTCATATGTTGAGGGACAGAAAGAAGAGTCAGTAC   1451

Qy   1381 caggagtacaggaggagtgcagagaacggagagaaagaaagtcagtac   1440
          |||||||||||||||||||||||||||||||||||||||||||
Db   1452 CAGGAGTACAGGAGGACTGCCAGGACGGGCCCAGCCCAGCGCCGCCCAGCGCCCG   1511

Qy   1441 ccagaaaatcatctccttggaacaagggtctgccatcccgatcccgatcgaaatgtcagt   1500
          ||||||||||||||||||||||||||||||||||||||||||
Db   1512 CCAGAAATCATCTTCCTTGGAACACGGGTTCGCCATCCCGATGAAGAATTCGAAAATGTCAGT   1571
```

```
Qy   1501 gccacaactgtcaacataagccccgacacgtcttctgctactggactggtggtgttgtgagggcaca   1560
          ||||||||||||||||||||||||||||||||||||||||||
Db   1572 GCCACAACTGTTCAACATAAGCCCCGACACACGTTCTCGCTACTGGACTGGTGGTGCTGCTACTGGACTGGTGGGCACA   1631

Qy   1561 tttgggcagctgtgccgttcattacggagacccacacacagtggacaggtgtcctgggcacctggct   1620
          ||||||||||||||||||||||||||||||||||||||||||
Db   1632 TTTGGGCACCTGTGTCCGTGCTGTCATTACGGAGACCACACAGTGGGACAGGGTCCTGGGTTGGCT   1691

Qy   1621 gctgtgttttgtgtccacccagcagatcaacacacagacacggcttggccaagtcacctgctg   1680
          ||||||||||||||||||||||||||||||||||||||||||
Db   1692 GCTGTGTTTTGTGTCCCACCCAGCAGATCAACACACAGACACGGGCTTGGCCAAGTCACCTGCTG   1751

Qy   1681 caggagaacgcgtcccttgccctttggcatctcatcttcttgggaaggggctccacccctttgtc   1740
          ||||||||||||||||||||||||||||||||||||||||||
Db   1752 CAGGAGAACGCGCGTCCCTTGGCCTATCTTCTTTGCTGGGATGTCACCCCTTTGCTAGGCGGTTGCC   1811

Qy   1741 cccaaccagctcaaagctcggcctccagcaagtaccacacacaaccagtgccagtacctggtcac   1800
          ||||||||||||||||||||||||||||||||||||||||||
Db   1812 CCCAAACCAGCTCAAAGCCTTGGCTCCAGCAGTACTACACCACACAGGCGGCTTGGCCAAA   1871

Qy   1801 cacatcagttatgattccctgccaaatgccttcaggaaggggctgaagatctccagctcagca   1860
          ||||||||||||||||||||||||||||||||||||||||||
Db   1872 CACATCAGTTATGATTCCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGCTCCAGCGCCTGCA   1931

Qy   1861 gtggaaaaagattgatccagcttcgcgttgccaccgttgggaaatggtttggggaactggtgt   1920
          ||||||||||||||||||||||||||||||||||||||||||
Db   1932 GTGGAAAGAATTGAATCAGTTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGA   1991

Qy   1921 ctgttgtcgcggcacctcgaagcactgggctcccgtcgtgtgcgtcggtgtcacacctctggtctggggaaa   1980
          ||||||||||||||||||||||||||||||||||||||||||
Db   1992 CTGTTGTGGCGGCACTCCACAAGCGTGGGCTTTGGCTCGTGTGTGTGTCGGGAAA   2051

Qy   1981 gtggtctattccgggacaccaccctgccctgcgagactcggtcggtatgtgggaagagatgcc   2040
          ||||||||||||||||||||||||||||||||||||||||||
Db   2052 ATGCTGAACCACTTCAGCCACGAGCCTATGCCAAGGTCCCCCGCTTCAGCCCCAACTTCAGC   2111

Qy   2041 acctcctgatacatgaagcccacccctggaaggatggttttgggaacctgt   2100
          ||||||||||||||||||||||||||||||||||||||||||
Db   2112 ACCCTCCTGATACATGAAGCCCACCCCTGGAAGGATGTTTTGGGAACAATG   2171

Qy   2101 acacacagcacacgtccccaaccatcacagcgtttggcttggggagttcatt   2160
          ||||||||||||||||||||||||||||||||||||||||||
Db   2172 ACACACAGCACACGTCCCCAAGCTCCCAAGCATGGGGGATGGAACCGGAGTTCATT   2231

Qy   2161 atgctgaaccacttcagccacgagctatgccaaggtccccccaacttcagc   2220
          ||||||||||||||||||||||||||||||||||||||||||
Db   2232 ATGCTGAACCACTTCAGCCACGAGCCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAG   2291

Qy   2221 gaaaagtgggagttgtgccttttgaccacgcgacccagggttctgctttggcgacatctctgcttttga   2280
          ||||||||||||||||||||||||||||||||||||||||||
Db   2292 GAGAAAGTGGGAGTTGTGCCTTTTGACCACATGAAGGTGCCCTTGCCTTCCGAGTTTCAACAATG   2351

Qy   2281 cccaagctgattccccacctgaaagccctgtttgctgtggcacatcgaggagatggagggag   2340
          ||||||||||||||||||||||||||||||||||||||||||
Db   2352 CCCAAGCTGATTCCCCACTGAAAGCCCTGTTTCCTGTGGCCACATCGAGGAGATGGAGGAG   2411

Qy   2341 cgcaggagaagcggagactgcggcgtgcgggggcgccatgggtgccattatttactt   2400
          ||||||||||||||||||||||||||||||||||||||||||
Db   2412 CGCAGGAGAAAGCGGGACGCTGCGGCGGCCCCTCGTGTCCAGGGAGACTGGCA   2471

Qy   2401 gggcggcctggaggatggg   2418
          |||||||||||||||||||
Db   2472 GGGCGGCCTGGAGGATGGG   2489
```

```
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 2908)
AUTHORS     Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE       A candidate prostate cancer susceptibility gene at chromosome 17p
JOURNAL     Nat. Genet. 27 (2), 172-180 (2001)
PUBMED      11175785
REFERENCE   2  (bases 1 to 2908)
AUTHORS     Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE       Pan troglodytes ortholog of human HPC2/ELAC2
JOURNAL     Unpublished
REFERENCE   3  (bases 1 to 2908)
AUTHORS     Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE       Direct Submission
JOURNAL     Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     CDS             1..2481
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24920.1"
                     /db_xref="GI:10946497"
                     /translation="MWALCSLLRSAGRTMSOGRTISOAPARERPKDPLHLRTRE
                     KRGPSGCSGGPNTVLQVVAAGSRDSGAALYVFSEFRNYLFNGGEGIQRLMQEHKLKV
                     ARLDNIFLTRMIWSNVGGLSGMLTILKETGLPRCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVVQIPIHSEQRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPEHGVSQRRGVRDSSLVVAFICKLHLRGNFIVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAESVLVDSRYQQWMERGPDTQHLVLNENCASVHNLRSHKIQTQLNI
                     HPDIFPLLTTSFPCKKEGPTLSVPMVQGECLLKYQLRPRREWORDAIITCNPEEFIEA
                     LQLPNFQOSVQEYRRSAQDGRAPAEKRSQYPEIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLDCGEGTFGQLCRHYGDVDRVLGTLAAVFEVSHLHADHHTGLLNILLORERAL
                     ASLGKRTCDLEEFQTCLVRHCHHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     ISSLLRTCDLEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFN
                     LIHEATLEDGLEEEAAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFN
                     EKVGVAFDHMKVCFGDPATMPKLIPPLKALFAGDIEEMEERRREKRELRQVRAALLSRE
                     LAGGLEDGEPOQKRAHTEEPQAKKVRAQ"
BASE COUNT      712 a    788 c    819 g    589 t
ORIGIN
Qy           1 atgtgggcgcttgctcggctgcgctccggcgccggaacgcaccagtgtcgcaggggacgc 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Query Match        67.2%;   Score 1667;   DB 89;   Length 2908;
Best Local Similarity   99.4%;   Pred. No. 0;
Matches 2417;  Conservative   0;  Mismatches   15;  Indels   0;  Gaps   0;

Qy        1 atgtgggcgcttgctcggctgcgctccggcgccggaacgcaccagtgtcgcaggggacgc 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        1 ATGTGGGCGCTTGCTCGGCTGCGCTCCGGCGCCGGAACGCACCAGTGTCGCAGGGGACGC 60

Qy       61 accatatcgcaggcaccccgcccgcgcccgagcggcgccgcaaggagcggccgctgcgcc 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ACCATATCGCAGGCACCCCGCCCGCGCCCGAGCGGCGCCGCAAGGAGCGGCCGCTGCGCC 120

Qy      121 cgcacgcgagagaagcgcgggaccgtcgggtgtgctccgggcggcccaaaacaccgtgctg 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CGCACGCGAGAGAAGCGCGGGACCGTCGGGTGTGCTCCGGGCGGCCCAAAACACCGTGCTG 180

Qy      181 caggtggttggcagcggtagccgggactcggcgcgcgctctaggtcttctccgggttc 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CAGGTGGTTGGCAGCGGTAGCCGGGACTCGGCGCGCGCTCTAGGTCTTCTCCAGTTC 240

Qy      241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaagagcacaagtta 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAAGAGCACAAGTTA 300

Qy      301 aaggttgctctgccctggacaacatattccctgaacacgaatgactgtctaatgttgggg 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AAGGTTGCTCTGCCCTGGACAACATATTCCCTGAACACGAATGACTGGTCTAATGTTGGGGC 360

Qy      361 ttaagtggaatgattcttactttaaggaaaccgggcttccaaagtgtgtacttttctgga 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 TTAAGTGGAATGATTCTTACTTTAAGGAAACCGGGCTTCCAAAGTGTGTACTTTTCTGGA 420

Qy      421 cctcacaaactggaaaaaatacctcgaagcaatcaaaatattttctggttccattgaaaga 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 CCTCACAAACTGGAAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTTCCATTGAAAGGA 480

Qy      481 atagaactgctgtcgcggcccactcctgccccagaaatacgaaggtgaaacatgacagtt 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ATAGAACTGCTGTCGCGGCCCACTCCTGCCCCACACTCTGCCCCAGAATACGAGGTGAAACCATGACAGTT 540

Qy      541 taccagatccccatcacacagtgaacagaagaggagggaaagcaccccaaccatggcagagtcca 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 TACCAGATCCCCATCACACAGTGAACAGAAGAGGAGGGAAAGCACCCCAACCATGGCAGAGTCCA 600

Qy      601 gaaaggccctccaagcaggctcagtccagagggatctcagactccgagtcgaataaaat 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GAAAGGCCCTCCAAGCAGGCTCAGTCCAGAGGGATCTCAGACTCCGAGTCAAATGAAAAT 660

Qy      661 gagccacaccttccacatggtgttagccagagaaggggtcagggagactcttccctggtc 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGGGGTCAGGGACTCTTCCCTGGTC 720

Qy      721 gtagcttcatctgtaagcttcacttaaagaagaagaaaacttcttggtgtccaaagcaaag 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 GTAGCTTCATCTGTAAGCTTCACTTAAAGAAGAAGAAACTTCTTGGTGTCCAAAGCAAAG 780

Qy      781 gagatggggcccccagttggacagctgcatcgctcccatctgctgctgtcaaggac 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 GAGATGGGGCCCCCAGTTGGACAGCTGCATCGCTCCCATCTGCTGCTGTCAAGGAC 840

Qy      841 gggaaaagcatcactcatgaagagagagagagattttggctgaagactgtgtactcctcca 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 GGGAAAAGCATCACTCATGAAGAGAGAGATTTTGGCTGAAGACTGTGTACTCCTCCA 900

Qy      901 gatcctggttgctgcttttgtggtggtagaatgtccagatgaaaagcttcattcaaaccaatc 960
            ||||||||||||||||||||||||||||||||||||||||||||||
Db      901 GATCCTGGTTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAAGCTTCATCAACCCATC 960

Qy      961 tgtgaagaatgccacctttcagaggtaccaagaggcagtgcccgtgggccttggtg 1020
            ||||||||||||||||||||||||||||||||||||||||||||||
Db      961 TGTGAGAAATGCCACCTTTCAGAGGTACCAAGAGGCAGTGCCCGTGGGCCTTGGTG 1020

Qy     1021 gttcacatggccccagcaatctgtcttgttggacagcagttaccagccagtggatggaagagg 1080
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 GTTCACATGGCCCCAGCAATCTGTCTTGTTGGACAGCAGTTACCAGCAGTTGGATGGAAGAGG 1080

Qy     1081 tttgggcctgacacccagcacttggtccctgaatgaagaactgtgcctcagttcacaacctt 1140
            ||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 TTTGGGCCCTGACACCCAGCACTTGGTCCCTGAATGAAGAACTGTGCCTCAGTTCACAACCTT 1140
```

```
Qy  1141 cgcagccacaagaagattcaaacccagctcaacccggacactcttccccctgctc 1200
         |||||||||||| ||||||||||||||||||||||||||||| |||||||||||||
Db  1141 CGCAGCCACAAGAGATTCAAACCCAGCTCAACCCGGACACATCTTCCCCTGCTC 1200

Qy  1201 accagttctccgctgtaagaagaggagggccccacccctcagtgtgccatggttcaggtgaa 1260
         ||||||| ||||| |||||||||||||||||| |||||||||||||||||||||||||||||
Db  1201 ACCAGTTTCCCCTGTAAGGACAAGAGGGCCCACCCTCAGTTGTGCCATGGTTCAGGTGAA 1260

Qy  1261 tgcctcctcaagtaccagctccgtcctccagaggaggagggatgtccattattact 1320
         |||||||||||||||||||||||||||||||||||||| ||||| |||||||||||
Db  1261 TGCCTCCTCAAGTACCAGCTCCGTCCTCCAGGAGGAGGGATGTCCAATTATTACT 1320

Qy  1321 tgcaatcctgagggaattcatagttgagggcgctgcagcttcccaacttccagcagaggtg 1380
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TGCAATCCTGAGGGAATTCATTAATTGAGGGCGCTGCAGCTTCCCAACTTCCAGCAGAGTGTG 1380

Qy  1381 caggagtacaggaaggaggagtgcgcaggcaggcagacggccccgagggagaaagaaagtcagtac 1440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CAGGAGTACAGGAGGAGTGCGCAGGCAGGCAGACGGCCCCGAGAGAGAAAGAAAGTCAGTAC 1440

Qy  1441 ccagaaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CCAGAAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1500

Qy  1501 gccacacttgtcaaacataagcccgacaacgtctctcgtactggacactgtgtgaggggaaca 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GCCACACTTGTCAACATAAGCCCGACAACGTCTCTCGTTACTGGACACTGTGTGAGGGGCACG 1560

Qy  1561 tttggcagctgtgcgtgatcattacggagacaggtggacagggtctgggcaaccctggct 1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 TTTGGGCAGCTGTGCGTGATCATTACGGAGACAGGTGGACAGGGTCCTGGGCAACCCTGGCT 1620

Qy  1621 gctgtgtttgtgtccacctgcacgcagatcaccacacggggttgccaagtatctctgctg 1680
         |||||| ||||||||||||||||||||| |||||||||||| |||||||||||||||||
Db  1621 GCTGTGTTTGTGTCCCACCTGCCAGATCACCACACGGGCCTTGCCAAGTATCTCTGCTG 1680

Qy  1681 cagagagaacgccctggcatcttgggaagcgcttcacccttgctcggtggttgcc 1740
         ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 CAGAGAGAACGAGCCGTTGGCAGTCTTTGGGAAAGCCCTTCACCCTTTGCTGGTGGTTGCC 1740

Qy  1741 cccaaccaagctcaaagctctgcctgccagtagccaaccaccagtgccaggaggtcctcgcac 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 CCCAACCAAGCTCAAAAGCCTGGCTGCCAGCAGTACCACACAGTGCCAGGAGGTCCTGGAA 1800

Qy  1801 cacatcagtgtgattcctgccaaatgccttcagagagctcctgagatctccagtcctgca 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGCTGAGATCTCCAGTCCTGCA 1860

Qy  1861 gtggaaagattgatcagttcgctgttgcggaacaatgtgattctggaagaggttcagacctgt 1920
         |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  1861 GTGGAAAGATTGATCAGTTCGCTGTTGCGGAACATTGTGATTTGGAAGAGTTTCAGACCTGT 1920

Qy  1921 ctggtgcggcaactgcaaagcatgcgtttggctgtgctgtgcgctgtgtgtcaacctctggctgaaa 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CTGGTGCGGCACTGCAAGAGCATGCGTTTGGCTGTTGCTGCCAGGTCCACACCTCTGGCTGGAAA 1980

Qy  1981 gtggtctattccgggacaccacatgcccctcggagggctgggctcgtggggaaagatgcc 2040
         |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  1981 GTGGTCTATTCCGGGGCACCACATGCCCCTCGGAGGGCTGGGGTTCGGTGGGGAAAGATGCC 2040

Qy  2041 aacctcctgatacatgaagccatccggaagatggttctgggaagagaacagcagtgagaaag 2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 ACCTTCCTGATACATGAAGCCACCCGGAAGATGGTTTGGGAAGAGGAACAGCAGTGGAAAG 2100

Qy  2101 acaacagcacaacgtcccaagccatcagcgtgggatgcgaaacgcggagttcatt 2160
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  2101 ACACACAGCACAACGTCCCAAGCCATCAGCGTTGGGATGCGATGACGCCGAGTTCATT 2160

Qy  2161 atgctgaaccacttcagcccgcgctatgccaaggtccccctccttcagccccaacttcagc 2220
         ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  2161 ATGCTGAACCACTTCAGCCAGCGCTTATGCCAAGGTCCCCCTCTTCAGCCCCAAACTTCAAC 2220

Qy  2221 gagaaagtgggagttgccttttgaccactgaaggtctgctttggagagactttcaacaatg 2280
         ||||| |||||||||||||||||||||||| |||||||||||||||||||||| |||||||
Db  2221 GAGAAAGTGGGAGTTGCCTTTTGACCACATGAAGGTCTGCTTTGGAGAGACTTTGCAACAATG 2280

Qy  2281 cccaagctgaattcccccactgaaaagcccctgtttgctcgtggcgaacatcgaggagagtggagag 2340
         ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  2281 CCCAAGCTGAATTCCCCCACTGAAAAGCCCCTGTTTGCTCGGGACAATCGAGGAGAGTGGAGAG 2340

Qy  2341 cgcaggggagagcgggagagctgcggcaggtgcggcagggccctccctctgtccaggagaactggca 2400
         |||||||||||| |||||||| |||||||||| ||||| ||||||| ||||||| ||||||||
Db  2341 CGCAGGGGAGAAAGCCGGGAGCTGCGGCAGTGTGCGGCGGGCCCTCCGTGTCCAGGGAGCTGGCA 2400

Qy  2401 gggcgcctggagatggggggagcctcagcagaa 2432
         || ||| ||||| |||||||| ||| |||||||
Db  2401 GGCGGGCCTGGGAGGATGGGGGGGAGCCTTCAGCAGAA 2432
```

RESULT   6

LOCUS   AF308694   2893 bp   mRNA   PRI   27-FEB-2001
DEFINITION Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION   AF308694
VERSION   AF308694.1   GI:10946488
KEYWORDS   .
SOURCE   gorilla.
  ORGANISM Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1 (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE   A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED   11175785
REFERENCE   2 (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE   Gorilla gorilla ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES   Location/Qualifiers
     source   1..2893
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
     gene   1..2893
                     /gene="ELAC2"
     CDS   1..2481

/gene="ELAC2"
/note="Similar to Homo sapiens HPC2/ELAC2"
/codon_start=1
/product="ELAC2"
/protein_id="AAG24916.1"
/db_xref="GI:1094689"
/translation="MWAICSLLRSAAGRHMSQSRTISQAPARRERPRKDPLRLRTRE
KRPGSGCSGGPNTVYLQVVAAGSRDSGAALYVPSEFNRYLFNCGEGVQRLMQEHKLKV
VRLDNIFLTRMHNSNVGGLSGMIILTLKETGLPKCVLSGPPQLEKYLEATKIFSGPLKG
IELAVRHSAPEYEDEDTMTVYQLPIHSEQRGRHQPWQSPERPLSRLSPERSDSDSSN
ENPPHLPHGVSQPRGVRDSSLVVAFICKHLHKRGNFLVLKAKEMGLPVGTAAIAPIIA
AVRDGKSITHEGRRILAEELCTPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
APVALVVHMAPSSVLVDSRYQOWMERGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
HDIFPLLTSFPCKKEGPTLSVPMVQGECLLKYOLRPREWQRDAIICNPEEFIVEA
LQLPNFQQSVQEYRRSVQDVPAPAEKRSQYPEIFLGTCSAIPMKIRNVSATLVNISP
DYSLLDCGEGTPGOLCRHYGDGVDRVLGTLAAVFVSHLHADHHTGLLNILLOREQAL
ASGKPLHPLLVVAPSQLKAWLQOYHNQCOEVLHHISMIPAKCLQEGAEISSPAVERL
ISSLRTCDLEEFQTCLVRRCKHAFGCALVHTSGWKKVVYSGDTMPCEALVRMGKDATL
LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHSRQYAKVPLFSPNFN
EKVGVAFDHMKVCFGDFPTMPKLIPPLRALFAGDIEEMEERREKRELRQVRAALLSGE
LAGDLEDGEPQORRAHTEPQAKKVRAQ"

BASE COUNT   705 a   787 c   815 g   586 t
ORIGIN

Query Match   51.0%;   Score 1266;   DB 89;   Length 2893;
Best Local Similarity   99.1%;   Pred. No. 0;   Mismatches 22;   Indels 0;   Gaps 0;
Matches 2366;   Conservative 0;   Mismatches 22;   Indels 0;   Gaps 0;

Qy   1 atgtgggcgcttgctccgtgtcgtcggtgccggacgacagactgccaccatgtcgcaggga gacgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||-||
Db   1 ATGTGGGCGGCTTTGCTCCGTGTCGTCGGTGCCGGGACGACACCATGTCGCAGGGACGC 60

Qy   61 accatatgcaggcacccgcgcccgcccgagcgcgccgaccgctgggacctg 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCATATGCCAGGCACCCCGCCCGCCCGCGCAAGGACCCGCTGCGCGACCTG 120

Qy   121 cgcacgcgagagaagcgcggaccgtcgggggtgtctccggggtctccg 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGCACGCGAGAGAAGCGCGGGACCGTCGGGGGTGTCTCCGGGGGTACCTG 180

Qy   181 caggtggtgcaggggggtagccggggcgccggcgctctcacgtctcctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGGTGGTGCAGGGGGGTAGCCGGGACTCGGGCGCCGCCCTCTACGTCTTCTCCGAGTTC 240

Qy   241 aaccggtatctctctcaactgtgagaaagccgttcaggagaactcatgcaggagtta 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AACCGGTATCTCTTCAACTGTGGAGAAGCGGTTCAGGAGACTCATGCACAAGTTA 300

Qy   301 aaggttgctcgcctggaacacacatattcctgacacgaatgcaactggtctaaatgttgggggc 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGGTTGTTGCCCTGGACAACATATTCCTGACACGAATGCACAATAGTCACTAC 360

Qy   361 ttaagtggaatgattctttaaaggaaaccgggcttccaaagtgtgtacttctttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTAAGTGGAATGATTCTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 420

Qy   421 cctccacaactggaaaaatacctcgaagcaatcaaaaatatttctgttccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCTCCACCACTGGAAAAAAATACCTCGAAGCAATCAAAAATATTTCTGTTCCATTGAAAGGA 480

Qy   481 atagaactggctgtggcgccccactctgcccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATAGAACTGGCTGTGGCGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 540

Qy   541 taccagatccccatacacagtgacagagaggagggaagcaccaacatggcacagagtcca 600
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TACCAGATCCCCATACACAGTGACACAGAGAGGAGGGAAGCCACCAACATGGCACAGAGTCCA 600

Qy   601 gaaaggcctcagcaggctcagtccgagcggatcttctagagactcgagacatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAAAGGCCTCTCAGCAGGCTCAGTCCGAGCGGATCTTCAGACTCCGAGTCGGAATCAAAAAT 660

Qy   661 gagccacacttccacatgtggtagccaggaccaagaggggtcaggggactcttcctggtc 720
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GAGCCACACCTTCCACATGGTGTTAGCCCAGAGAGAGAAGAGGGGTCAGGGGACTCTTCCTGGTC 720

Qy   721 gtagcttcatcgtaagcttcacttaaagaggaggaaacttcttggtgctcaaaagcaaag 780
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGGAGGAAACTTCTTGGTGCTCAAAAGCAAAG 780

Qy   781 gagatggggcctccagctgccatcgcctccacatcattgctgctcggtctcaaggac 840
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GAGATGGGGCCTCCCAGCTGCCATCGCCTCCCACATCATTGCCTGCTCGGTCTCAAGGAC 840

Qy   841 gggaaaagcatcaactcatgaaggaagagaagagatttggctagacagaagaccctcca 900
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GGGAAAAGCATCAACTCATGAAGGAAGAGAAGAGATTTGGCTAGACAGAAGACCTCCCCA 900

Qy   901 gatcctggctgctgttttggttggttgaagatgtccagatgaaagcttcatcaaccatc 960
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GATCCTGGTTGCTGTTTTTGCTTTTGCGGTGGTTGAAGATGTCCAGATGAAAGCTTCATCAAGGAC 960

Qy   961 tgtgagaatgccaccttttcaggagtaccaagcaaggaaaagcagatgcccccgtggcctggtg 1020
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TGTGAGAATGCCACCTTTTCAGGAGTACCAAGCAAGGAAAAGCAGATGCCCCCGTGGCCTGGTG 1020

Qy   1021 gttcacatggcccccagacatctgtgcttgtggacagacaggtacccagacagtggatggagagg 1080
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCCAGAATCTGTGCTTGTGGACAGACAGGTACCCAGCAGTGGATGGAGAGG 1080

Qy   1081 tttgggcctgacaccccagacacttgtcctgaatgaatgagaaactgcccagtcaacactt 1140
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGGCCTGACACCCCAGACACTTGGTCCTGAATGAGAGAACTGGCCCAGTCCAACACTT 1140

Qy   1141 cgcagcgcacaagagattcaaacccagtctcaaacctcaatccaaccggacaatcttccccctgctc 1200
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAGCGCACAAGAGATTCAAAACCCAGTCTCAAACCTCATCCAACCGGACATCTTCCCCGCTC 1200

Qy   1201 accagttccgctgtaagaaggagggcccccaccctcaagtgtgccatggttcaagggtaa 1260
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAGTTTCCGCTGTAAGAAGGAGGGGCCCCCACCCTCAAGTGTCCCATGGTTCAGGGTGAA 1260

Qy   1261 tgcctcctcaagtaccacagctccgtcccaggagggaggagggggatgccaattattact 1320
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGCCTCCTCAAGTACCACAGCTCCGTCCCAGGAGGGAGGGGGATGCCAATTATCACT 1320

Qy   1321 tgcaatcctgaggaattcatagttgagtgcgtcagctgttcccaacttccagagaaggtg 1380
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGCAAATCCTGAGGAATTCATATGGTTGAGTGCCGTCAGCTGTTCCCAACTTCCAGAGAGTG 1380

Qy   1381 caggagtacaggaggagtgcgcagagacggcccagcagagaggaaaagaaagtcagtac 1440
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGGAGTACAGGAGGAGTGTGCAGGAGACGGTCCATGGTTCAGGTTGAA 1440

Qy   1441 ccagaaatcatctccttggaacagggtctgccatcccgatgaagagattccagt 1500
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCCATGAAGAGATTCGAAATGTCAGT 1500

Qy   1501 gccacacttgtcaaacataagcccgacacgtctctcgttgctacgtggtgtgaggggcaca 1560
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCACACTTGTCAAACATAAGCCCGACACGTCTCTCGTTGCTACGTGGTGTGAGGGGCACG 1560

Qy   1561 tttgggcagctgtccgtcattacggagaccaggtcctgggcacccctggct 1620
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGGGCAGCTGTCCGTCGCTCATTACGGAGACCAGGTCCTGGGCACCCCTGGGCT 1620

Qy   1621 gctggtgttgtgtccaacctgcacgagatacgtggcaagtatctctgtg 1680
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCTGTGGTTGTGTGTCCAACCTGCACGAGATACCACCAGTCCTTGCTAAATATCTTGCTG 1680

Qy   1681 caggagaaacggcgcctggcatcttgggaaaagccgttcaccccttttgctggttggtcc 1740
          ||||||| --|||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGGAGAACAACAAGCCCTTGGCATCTTTGGGAAAAGCCCTTCACCCTTTGCTGGTTGGTGCC 1740

Qy   1741 cccaaccaagctcaaagcctggctccagacagtaccacacaaccagtgccaggagggtcac 1800

```
Db   1741  CCCAGGCAGCTCCAAAGCCTGGCTTCCCAGCAGTACCAACCAGTGCCAGGAGGTCCTGCAC
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1801  cacatcagtatgatcctgccaaatgccttcaggaagggctgagatctccagtcctgca  1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  CACATCAGTATGATCCTGCCAAATGCCTTCAGGAAGGGCTGAGATCTCCAGTCCTGCA  1860

Qy   1861  gtggaaagattgatcaattcgctgttgcgaacatgtgattggaagagttcagacctgt  1920
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   1861  GTGGAAAGATTGATCACTTCGCTGTTGCGAACATGTGATTGGAAGAGTTTCAGACCTGT  1920

Qy   1921  ctggtgcggcactgcaagcatcgctgtttggctgtgctcgtcgtgtgcaacctctggctggaa  1980
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1921  CTGGTGCGGCACTGCAAGCATCGCTGTTTGGCTGTGCTCGTCGTGTCACACCTCTGGCTGGAA  1980

Qy   1981  gtggtctattccggggaccaccatgcctggttccgggatgggagaaagatgcc  2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  GTGGTCTATTCCGGGGACCACCATTGCCTGGTTCCGGGATGGGAGAAAGATGCC  2040

Qy   2041  accctcctgatacataatgaagccacccctggaagatggtttttggaagagaagacagtggaaag  2100
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   2041  ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTTGGGAAGAGAAGCAGTGGAAAG  2100

Qy   2101  acacacagcacaacgtcccaagccatcaagcgtggggatgcggcggagttcatt  2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  ACACACAGCACAACGTCCCAAGCCATCAAGCGTGGGGATGACGCGGAGTTCATT  2160

Qy   2161  atgctgaaccacttcagccagcgctatgccaagtgtccccctcttcagccacccccaacttcagc  2220
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   2161  ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGTGTCCCCCTCTTCAGCCCAACTTCAAC  2220

Qy   2221  gagaaagatgggagttgcctttgacccacaatgatctgcttctgctggagacttccaacaatg  2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  GAGAAAGTGGGAGTTGCCTTTGACCATGGCCTTTGCTGGAGACTTCCAACAATG  2280

Qy   2281  cccaagctgattccccactgaaagccctgtttgctggcgacatcgaggagaggatggaggag  2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  CCCAAGCTGATTCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGAGGATGGAGGAG  2340

Qy   2341  cgcaggagaagcggagctgcggagaggtgcgggggcggcccctcctgtgtcc  2388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  CGCAGGAGAAGCGGAGCTGCGGAGAGGTGCGGGGGCGGCCCCTCCTGTGTCC  2388
```

RESULT   7
LOCUS   AC005277   118788 bp   DNA   PRI   23-JUL-1998
DEFINITION Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
ACCESSION   AC005277
VERSION   AC005277.1 GI:3337311
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 118788)
AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens chromosome 17, clone hRPK.597_M_12
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 118788)
AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky},J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 118788)
AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky},J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jul 23, 1998 this sequence version replaced gi:3335015.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Only the first 118.8 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC005274 (WICGR project
            L350).
FEATURES   Location/Qualifiers
     source   1..118788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.597_M_12"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   795..884
                     /rpt_family="MIR"
     repeat_region   3822..4007
                     /rpt_family="MLT1A1"
     repeat_region   complement(3478..3821)
                     /rpt_family="L2"
     repeat_region   complement(932..1000)
                     /rpt_family="MIR"
     repeat_region   2044..2076
                     /rpt_family="AT_rich"
     repeat_region   complement(4015..4281)
                     /rpt_family="MER89"
     repeat_region   2237..2713
                     /rpt_family="MER34"
     repeat_region   2718..3477
                     /rpt_family="L2"
     repeat_region   complement(4609..4895)
                     /rpt_family="AluSx"
                     /rpt_family="MER89"
     repeat_region   complement(4282..4321)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(4896..5186)
                     /rpt_family="MER9"
     repeat_region   complement(4328..4608)
                     /rpt_family="AluSp"
     repeat_region   5219..5479
                     /rpt_family="L2"
     repeat_region   complement(5481..5740)
                     /rpt_family="MER34"
     repeat_region   complement(5963..6275)
                     /rpt_family="AluSp"
     repeat_region   complement(6411..6494)
                     /rpt_family="MERSA"
     repeat_region   6830..7013
                     /rpt_family="MER5A"

```
repeat_region    /rpt_family="MLT1C"
                 7014..7318
repeat_region    /rpt_family="AluYb8"
                 7319..7581
repeat_region    /rpt_family="MLT1C"
                 7721..8032
repeat_region    /rpt_family="AluJb"
                 8039..8333
repeat_region    /rpt_family="AluSq"
                 9431..9460
repeat_region    /rpt_family="AT_rich"
                 9662..9829
repeat_region    /rpt_family="MIR"
                 complement(11302..11728)
repeat_region    /rpt_family="MER39"
                 complement(11824..11884)
repeat_region    /rpt_family="LTR16C"
                 complement(11928..12108)
repeat_region    /rpt_family="LTR16C"
                 12453..12725
repeat_region    /rpt_family="AluJb"
                 13377..13785
repeat_region    /rpt_family="MSTB"
                 14118..14153
repeat_region    /rpt_family="(CA)n"
                 complement(15320..15620)
repeat_region    /rpt_family="AluJo"
                 15683..15759
repeat_region    /rpt_family="(CA)n"
                 complement(15838..15893)
repeat_region    /rpt_family="MER91A"
                 complement(15915..16080)
repeat_region    /rpt_family="MER91A"
                 complement(16474..16544)
repeat_region    /rpt_family="L2"
                 17907..17920
repeat_region    /rpt_family="L1ME"
                 18271..18311
repeat_region    /rpt_family="L2"
                 complement(18368..18529)
repeat_region    /rpt_family="AluJb"
                 complement(18530..18560)
repeat_region    /rpt_family="(CA)n"
                 complement(18561..18681)
repeat_region    /rpt_family="AluJb"
                 18760..19065
repeat_region    /rpt_family="AluSg"
                 complement(19638..19896)
repeat_region    /rpt_family="MIR"
                 complement(19971..20260)
repeat_region    /rpt_family="L1ME3A"
                 20263..20552
repeat_region    /rpt_family="MER58B"
                 complement(20553..20816)
repeat_region    /rpt_family="L1ME3A"
                 complement(20821..21069)
repeat_region    /rpt_family="L1M4"
                 complement(21071..21100)
repeat_region    /rpt_family="(TAAA)n"
                 complement(21101..21338)
repeat_region    /rpt_family="AluSx"
                 complement(21348..21598)
repeat_region    /rpt_family="L1M4"
                 21799..22104
repeat_region    /rpt_family="AluJo"
                 22151..22562
repeat_region    /rpt_family="MSTA"
                 23978..24287
repeat_region    /rpt_family="AluSq"
                 24852..25199
repeat_region    /rpt_family="L2"
                 complement(25697..25985)
repeat_region    /rpt_family="MLT1D"
```

```
unsure           25816..25895
                 /note="Single-stranded coverage."
repeat_region    complement(25994..26099)
                 /rpt_family="purine-rich"
repeat_region    complement(26122..26291)
                 /rpt_family="MLT1D"
repeat_region    26312..26337
                 /rpt_family="(CAAAA)n"
repeat_region    26456..26480
                 /rpt_family="(CAAA)n"
repeat_region    complement(26576..26781)
                 /rpt_family="(GGGAA)n"
repeat_region    complement(26912..26990)
                 /rpt_family="MER6B"
repeat_region    26991..27290
                 /rpt_family="MER21B"
repeat_region    complement(27291..27975)
                 /rpt_family="AluSx"
repeat_region    28034..28197
                 /rpt_family="MER21B"
repeat_region    28266..28795
                 /rpt_family="LTR28"
repeat_region    28843..28916
                 /rpt_family="LTR28"
repeat_region    28920..29005
                 /rpt_family="LTR28"
repeat_region    29140..29233
                 /rpt_family="MIR"

Query Match      10.0%;   Score 247;   DB 85;   Length 118788;
Best Local Similarity   100.0%;   Pred. No. 2.6e-131;
Matches    247;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy      1 atgtggcgcttctcgctgctgcggtccgcggccggacggcaccatgtcgcaggggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78562 atgtggcgcttctcgctgctgcggtccgcggccggacggcaccatgtcgcaggggacgc 78621

Qy     61 accaatatcgcaggcacccgcccgcccgccgagcgggcccgcgcaaggagaccggctgcgggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78622 accaatatcgcaggcacccgcccgcccgccgagcgggcccgcgcaaggagaccggctgcgggcacctg 78681

Qy    121 cgcacgcgagagaagcgcggaaccgtcggggttgctccgcgcgggcccaaacacccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78682 cgcacgcgagagaagcgcggaaccgtcggggttgctccgcgcgggcccaaacacccgtgtacctg 78741

Qy    181 caggtggttgcacgggagtagcgggactcgggcgcgcccggcgcttcacgtcttctccggagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78742 caggtggttgcacgggagtagcgggactcgggcgcgcccggcgcttcacgtcttctccggagttc 78801

Qy    241 aaccggt 247
          |||||||
Db  78802 aaccggt 78808
```

REFERENCE   1 (bases 1 to 721)
AUTHORS   Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.,
            and Cannon-Albright,L.A.
TITLE   A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 740)
AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
TITLE   Direct Submission
JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES   Location/Qualifiers
     source   1..740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     exon   301..439
                     /number=17
     mRNA   <301..>439
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     CDS   <301..>439
                     /codon_start=2
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24440.1"
                     /db_xref="GI:10880931"
                     /translation="PDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADH
                     HT"
BASE COUNT   154 a   188 c   173 g   225 t
ORIGIN

Query Match   5.7%; Score 142; DB 89; Length 740;
Best Local Similarity   100.0%; Pred. No. 2.5e-70;
Matches 142; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1519 agcccgacacgtctctgctactggactgtggtgagggcacacattggggcaagctgtgccgt 1578
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   299 AGCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACACATTGGGGCAAGCTGTGCCGT 358

Qy   1579 cattacggagaccagtggacagggtcctgggcacacctggctgctgctgttgtgtcccac 1638
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   359 CATTACGGAGACCAGTGGACAGGGTCCTGGGCACACCTGGCTGCTGCTGTTGTGTCCCAC 418

Qy   1639 ctgcacgcagatcaccacaacgg 1660
           |||||||||||||||||||||||
Db   419 CTGCACGCAGATCACCACACACGG 440

RESULT 9
LOCUS   AF304371S1   721 bp   DNA   PRI   19-OCT-2000
DEFINITION Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 7 and partial cds.
ACCESSION   AF304371
VERSION   AF304371.1   GI:10880928
KEYWORDS   .
SEGMENT   1 of 2
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 721)

AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
TITLE   A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 721)
AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
TITLE   Direct Submission
JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES   Location/Qualifiers
     source   1..721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     exon   301..420
                     /number=7
     mRNA   <301..>420
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     CDS   <301..>420
                     /codon_start=3
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24439.1"
                     /db_xref="GI:10880930"
                     /translation="EQRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPH"
BASE COUNT   201 a   143 c   184 g   193 t
ORIGIN

Query Match   4.9%; Score 122; DB 89; Length 721;
Best Local Similarity   100.0%; Pred. No. 1.1e-58;
Matches 122; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   560 gtgaacacagaggagggaaagcacacaacatggcaagagagtcccagaaaagcctctaagcaggc 619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTGAACACAGAGGAGGGAAAGCACACAACCATGGCAAGAGAGTCCCAGAAAAGCCTCTAAGCAGGC 360

Qy   620 tcagtccagagccgatcttcaagactccgagtcgaatgaataaatgaggaccacaccttccacatg 679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TCAGTCCAGAGCCGATCTTCAGACTCCGAGTCGAATGAAAAATGAGCCACCACACCTTCCACATG 420

Qy   680 gt 681
           ||
Db   421 GT 422

RESULT 10
LOCUS   AF308696   2712 bp   mRNA   ROD   27-FEB-2001
DEFINITION Mus musculus ELAC2 mRNA, complete cds.
ACCESSION   AF308696
VERSION   AF308696.2   GI:11992378
KEYWORDS   .
SOURCE   house mouse.
ORGANISM   Mus musculus
            Mus musculus

```
REFERENCE    1  (bases 1 to 2712)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
             Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
             Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
             Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
             Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE      A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL    Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED    11175785
REFERENCE    2  (bases 1 to 2712)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
             Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
             Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
             Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
             Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
             Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
             Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
             Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
             and Cannon-Albright,L.A.
  TITLE      Mouse ortholog of human HPC2/ELAC2
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 2712)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
             Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
             Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
             Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
             Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
             Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
             Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
             Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
             and Cannon-Albright,L.A.
  JOURNAL    Direct Submission
             Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
             Lake City, UT 84103, USA
REFERENCE    4  (bases 1 to 2712)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
             Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
             Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
             Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
             Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
             Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
             Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
             Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
             and Cannon-Albright,L.A.
  JOURNAL    Direct Submission
             Submitted (26-DEC-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
             Lake City, UT 84103, USA
  REMARK     Sequence update by submitter
  COMMENT    On Dec 26, 2000 this sequence version replaced gi:10946492.
FEATURES             Location/Qualifiers
     source          1..2712
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="fetus"
     CDS             15..2510
                     /codon_start=1
                     /product="ELAC2"
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /protein_id="AAG24918.2"
                     /db_xref="GI:11923379"
                     /translation="MWALRSLLRPLGLRTMSQGSARPRPSKDPLRHLRTREKRGPGP
                     GGPNTVLVLQVVAAGGRDAGAALYVFSEYNRYLPKCVLSGPPOLEKYLEAIKIFSGP
                     LKGIELAVRPHTRMHWSNVGGLCGMILIFLKETGLPKCVLSGPPOLEKYLEAIKIFSGP
                     SAPEYKDETMTVQVPIHSERCKGKQPSQSPRTSPNRLSPKQSSDSSSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLLHLRGHGNEFLVLKAKELGLPVGTAIAPIAIA
                     GKSITYEGREIAAEELCTPPDGLVFIVVECPDEGFILPICENNCPSVHNLRSHKIQTOLSLIHPDI
                     LVVHIAPESVLIDSRYQOWMERGPDTOHLILNENCPSVHNLRSHKIQTOLSLIHPDI"
```

```
DEFINITION  Mus musculus putative prostate cancer susceptibility protein
            (Elac2) gene, complete cds, alternatively spliced.
ACCESSION   AF348157
VERSION     AF348157.1  GI:13540341
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (sites)
  AUTHORS    Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
             Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
             Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
             Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
             Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE      Mouse Elac2-containing genomic DNA
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 34593)
  AUTHORS    Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
             Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
             Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
             Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
             Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
             Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..34593
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     mRNA            join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16677,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27801,27930..28039,28553..28652,
                     28733..28853,29101..29179,29710..29854,30406..>30639)
                     /gene="Elac2"
                     /note="alternatively spliced 24 exon form"
                     /product="putative prostate cancer susceptibility protein"
     mRNA            join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
```

```
RESULT   11
AF348157    34593 bp    DNA                 ROD       04-APR-2001
LOCUS       AF348157
```

```
BASE COUNT    659 a    741 c    748 g    564 t
ORIGIN
```

```
Qy       445  gaagcaatcaaaatatttctggtccattgaaaggaatagaaactggc 491
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       435  GAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAAACTGGC 481
```

```
Query Match           1.9%;   Score 47;   DB 94;   Length 2712;
Best Local Similarity  100.0%;   Pred. No. 4.5e-15;
Matches   47;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```
FPQLTSFYSKEEGSTLSVPTVRGECLLKIYQLRPKREWQRDTTPDCNTDFIAEALELP
SFQESVEEYRKNVWQDENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
LLDCCEGTFGQLCRIYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
KPFQPLLVVAPTQLRAWLQQYHHCQEILHHVSMIPAKCLQGKGAEVSNTLERIISL
LETCDLEEFQTCLVRHCKHAFGCALVHSSGMKVVSGDTHWCEALVQMGKDATLLIHE
ATLEDGLEEEAVEKHSTTSQAINVGMRMNAGFIMLNHFSQRYAKIPLFSPDFNEKVG
IAFDHMKVCFGDFPVPVKLIPPLRALFAGDIEEMVERRKRELRLVRAALLITQQADSP
EDRRPQQKRAHTDEPHSPQSKKESVANTLGARV
```

```
Qy     445 gaagcaatcaaaatattttctggtccattgaaagggaataagaactgg 490
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   10969 GAAGCAATCAAAATATTTTTCTGGTCCATTGAAAGGGAATAGAACTGG 11014
```

RESULT 12
LOCUS   AC005545/c   43514 bp   DNA   PRI   03-SEP-1998
DEFINITION   Homo sapiens chromosome 19, cosmid R26634, complete sequence.
ACCESSION   AC005545
VERSION   AC005545.1   GI:3478638
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 43514)
AUTHORS   Lamerdin,J.E.; McCready,P.M., Skowronski,E., Adamson,A.W.,
             Burkhart-Schult,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S.,
             Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
             Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
             Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
             Coefield,J. Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
             Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
             Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE   Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
             serine protease gene cluster
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 43514)
AUTHORS   Lamerdin,J.E.
TITLE   Direct Submission
JOURNAL   Submitted (03-SEP-1998) Joint Genome Institute, Lawrence Livermore
             National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE   3 (bases 1 to 43514)
AUTHORS   Lamerdin,J.E.
TITLE   Direct Submission
JOURNAL   Submitted (27-AUG-1998) Joint Genome Institute, Lawrence Livermore
             National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT   Map and sequence oriented from p telomere to centromere. Cosmid
             R26634 overlaps cosmid R26660 (AC005328) to the left from bases 1
             to 4,683 of this accession, and overlaps cosmid F8682 (AC005257) to
             the right from bases 41,662 to 43,514. Additional chromosome 19 map
             and sequence information may be obtained at:
             http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES   Location/Qualifiers
     source   1..43514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19p13.3 between CDC34 and D19S342"
                     /cell_line="5HL2-B"
                     /clone_lib="LL19NC03 R chromosome 19-specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from hybrid 5HL2-B, which carries chromosome
                     19 as its only human chromosome."
     misc_feature   46..152
                     /note="DDS similarity to overlapping ESTs:
                     (46..152) AA431731 zw77g03.s1 Soares testis NHT Homo
                     sapiens cDNA clone 782260 3'; (391..340); 100%
                     identity.-(111..152) AA605572 af15c02.s1 Soares testis NHT
                     Homo sapiens cDNA clone 1031714 3'; (372..331); 100%
                     identity."
     misc_feature   complement(273..608)
                     /note="DDS similarity to overlapping ESTs:
                     (608..273) T05239 EST03128 Homo sapiens cDNA clone
                     HFBCT75. Score: 621 Identity: 325/337 (96%).-(718..308)
                     AA496009 zv72a03.s1 Soares total fetus Nb2HF8 9w Homo
                     sapiens cDNA clone 759148 3'; Score 755 Identity:
                     378/408 (97%)."
     misc_feature   378..707
                     /note="DDS similarity to overlapping ESTs:
```

gene   <8470..>31277
                /gene="Elac2"
                /note="alternatively spliced 25 exon form"
                /product="putative prostate cancer susceptibility protein"
gene   <8470..>31277
                /gene="Elac2"
                /note="24 exon form"
CDS   join(8470..8690,9077..9127,9266,9336,9948,10012,
                10957..11014,12096..12164,12933..13064,14902,14960,
                16609..16746,16818,17370..17482,18671,18766,
                21539..21677,21757..21842,23553,23671,24619,24715,
                27291..27429,27763..27801,27930..28039,28553,28652,
                28733..28853,29101..29179,29710..29854,30406..30624,
                31242..>31277))
                /gene="Elac2"
                /codon_start=1
                /product="putative prostate cancer susceptibility protein"
                /protein_id="AAK29421.1"
                /db_xref="GI:13540343"
                /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTEKRGPGP
                GGPNTVYLQVVAAGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGTIELAVRPH
                SAPEYKDETMTVYQVPIHSERRCGKQOPSQSPRTLSPKQSSDSGSAENGQCPPE
                DSSAGANRKAWGRDPSLVAFAVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                GKSITYEGRETAAEELCTPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                LVVHIAPESVLIDSRYQQWMERFGPDTQHILLINENCPSVHNLRSHKIQTOLSLIHPDI
                FPQLTSFYSKEEGSTLSVPTYRGECLLKYLQLRPKREWQRDTTLDCNTDEFIAEALELP
                SFQESVEEYRRNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                LLDCGEGTFGQLCRHYGQOIDRVLCSLTAVFSVSHLHADHHTGLLNILLQREHALASLG
                KPFQPLLVAPTQLRAWLQOYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                LETCDLEEFQTCLVRHCKHAPGCALVHSSSGWKVVVSGDTMPCEALVQMGKDATLLHE
                ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMILNHFSQRYAKIPLFSPDFNEKVG
                EDREPQOKRAHTDEPHSPQSKKESVANTLGARV"
CDS   join(8470..8690,9077..9127,9266,9336,9948,10012,
                10957..11014,12096..12164,12933..13064,14902,14960,
                16609..16746,16818,17370..17482,18671,18766,
                21539..21677,21757..21842,23553,23671,24619,24715,
                27291..27429,27763..27801,27930..28039,28553,28652,
                28733..28853,29101..29179,29710..29854,30406..30624,
                31242..>31277))
                /gene="Elac2"
                /codon_start=1
                /product="putative prostate cancer susceptibility protein"
                /protein_id="AAK29420.1"
                /db_xref="GI:13540342"
                /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTEKRGPGP
                GGPNTVYLQVVAAGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGTIELAVRPH
                SAPEYKDETMTVYQVPIHSERRCGKQOPSQSPRTLSPKQSSDSGSAENGQCPPE
                DSSAGANRKAWGRDPSLVAFAVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                GKSITYEGRETAAEELCTPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                LVHIAPESVLIDSRYQQWMERFGPDTQHILLINENCPSVHNLRSHKIQTOLSLIHPDI
                FPQLTSFYSKEEGSTLSVPTYRGECLLKYLQLRPKREWQRDTTLDCNTDEFIAEALELP
                SFQESVEEYRRNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                LLDCGEGTFGQLCRHYGQOIDRVLCSLTAVFSVSHLHADHHTGLLNILLQREHALASLG
                KPFQPLLVAPTQLRAWLQOYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                LETCDLEEFQTCLVRHCKHAPGCALVHSSSGWKVVVSGDTMPCEALVQMGKDATLLHE
                ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMILNHFSQRYAKIPLFSPDFNEKVG
                EDREPQOKRAHTDEPHSPQSKKESVANTLGARV"

BASE COUNT   8030 c   8142 g   9780 t
ORIGIN   8641 a
```

```
misc_feature    <378..468
                /note="Hypothetical human protein with no significant
                similarity to known proteins"
                /codon_start=2
                /evidence=not_experimental
                /product="R26660_2, partial CDS"
                /protein_id="AAC34213.1"
                /db_xref="GI:3478640"
                /translation="LVSPALRCLEPPHLANILIEDAAECLKQH"

CDS             <378..468
                /codon_start=2
                /product="R26660_2, partial CDS"
                /protein_id="AAC34213.1"
                /db_xref="GI:3478640"
                /translation="LVSPALRCLEPPHLANILIEDAAECLKQH"
```

AA431731 zw77g03.s1 Soares testis NHT Homo sapiens cDNA
clone 782260 3'; (339..11); 99% identity.-AA605572
af1sc02.s1 Soares testis NHT Homo sapiens cDNA clone
1031714 3'; (330..5); 99% identity.-AA448181 zw83c05.s1
Soares testis NHT Homo sapiens cDNA clone 782792 3';
(336..4); 99% identity.

```
misc_feature    1771..3383
                /note="DDS similarity to overlapping ESTs:
                (3391..3154) AA620776 ac14c10.s1 Stragtene Hela cell s3
                937216 Homo sapiens cDNA clone 856434 3'; similar to
TR:G163328 G163328 LEUKEMIA VIRUS CELL RECEPTOR. ;
                (262..1); 99% identity.-(1771..2449) AA167736 zq40d04.s1
                Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
                632167 3'; Score: 1207 Identity: 655/674
                (97%).-(2109..2638) AA843415 ak07f04.s1 Soares parathyroid
                tumor NbHPA Homo sapiens cDNA clone IMAGE:1405279 3';
                Score: 969 Identity: 516/547 (94%).-(2814..2510) AA364678
                EST75336 Pineal gland II;IV Homo sapiens cDNA clone
                582 Identity: 301/305 (98%).-(3200..2766) H73448
                yu02f02.r1 Homo sapiens cDNA clone 232635 5'. Score: 717
                Identity: 413/446 (92%).-(3358..2857) H28923 ym33c10.r1
                Homo sapiens cDNA clone 49874 5'. Score: 756 Identity:
                465/518 (89%).-(3383..2950) W69851 zd48c04.r1 Soares fetal
                heart NbHH19W Homo sapiens cDNA clone 343878 5'; similar to
                PIR:A45716 A45716 leukemia virus cell receptor ;
                Score: 744 Identity: 418/440 (95%). -Additional EST
                matches:
                AA617872, AA716429, W63772, AA305631, AA321888, AA371945,
                R18826, T33092, T30887, F01584 and many others"

mRNA            complement(2119..>39837)
                /product="Human AP-3 complex delta subunit mRNA"
                /note="Human AP-3 complex delta subunit mRNA"

CDS             complement(join(3296..3391,9810..9889,10209..10330,
                10996..11081,11255..11347,11830..12019,12408..12445,
                12802..12951,13983..14054,14459..14536,15248..15425,
                15871..15944,16342..16541,16661..16736,17330..17401,
                17728..17869,18345..18490,19724..19955,21985..22215,
                22226..22434,22857..23002,24481..24529,24953..25002,
                28275..28324,30213..30244,31625..31660,33594..33701,
                38134..38214,38850..38930,39742..>39837))
                /codon_start=1
                /product="AP-3 complex delta subunit, partial CDS"
                /protein_id="AAC34214.1"
                /db_xref="GI:3522925"
                /translation="AKYISQCIDEIKQELKQDNIAVKANAVCKLTYLQMLGYDISWAA
                FNIEVMSASKFTFKRIGYLAASQSFHEGTDVIMLTTNQIRKDLSSPSQYDTGVALTG
                LSCFVTPDLARDLANDIMTLMSHTKPYIRKKALIEPLTNLIHSTSAMSLIYECVNTVI
                AVILSLSSGMPNHSASIOLCVOKLRILIEDSDONLKYLGLLAMSKILKTHPKSVOSHK
                DLILQCLDDKDESIRLRALDLLYGMVSKKNLMEIVKLMTHVDKAEGTTYRDELLTKI
                IDICSQSNYQYITNFEWVISILVELTRLEGTRHGHLIAAQMLDVAIRVRAIRKFAVSQ
                MSALLDSAHLLASSTQRNGICEVLYAAAWICGEFSEHLLQEPHHTLEAMLRPRVTTLPG
                HIQAVYQNVVKLYASILQOKEQAGEAEGAQAVTQLMVDRLPQFVQSADLEVQERASC
                ILQLVKHIQKLQARDVPVAEEVSALFAGELNPVAPKLLEPLTNINEPLS
                DSESEDERPRAVFEHEEQRPKHRPSEADEELARRREARKQEQANNPFYIKSSPSPQ
                KRYQDTPGVEHIPVVOIDLSVPLKVPGLPMSDQYVKLEEERHRQKLEKDKRRKKRE
                KEKGKRRHSSLPTESDEDIAPAQQVDIYTEEMPENALPSDEDKDPNDYRALDIDLOKP
                DKPLADSEKLPIOKHRNTETSKSPEKDVPMVEKKSKPKKKEKHKEKERDKEKKKEK
                EKRAEEDLDFWLSTPPPAPAPAPAVPSTDECEDDAKTEAQGEDDAEGDQODKKSPKP
                KKKHRKEEERTRGKKSKSKKQPGSEEAAGEPVQNGAFEEEQLPPESSYSLLAENSY
                VKMTCDIRGSLOEDSQVTVAIVLENRSSSTILKGMELSVLDSLNARMARPQGSSVHDGV
                PVPQLPPGVSNEAQYVFTIQSIVMAQKLKGTLSFIAKNDEGATHEKLDFRLHFSCSS
```

```
CDS             complement(join(3296..3391,9810..9889,10209..10330,
                10996..11081,11255..11347,11830..12019,12408..12445,
                12802..12951,13983..14054,14459..14536,15248..15425,
                15871..15944,16342..16541,16661..16736,17330..17401,
                17728..17869,18345..18490,19724..19955,21985..22215,
                22226..22434,22857..23002,24481..24529,24953..25002,
                28275..28324,30213..30244,31625..31660,33594..33701,
                38134..38214,38850..38930,39742..>39837))
                /codon_start=1
                /product="delta-adaptin, partial CDS"
                /protein_id="AAC34212.1"
                /db_xref="GI:3478639"
                /function="similar to alpha and gamma adaptins"
                /note="AP-3 complex delta subunit"
                /function="vesicle coat component"
                /note="vesicle coat component, similar to alpha and gamma
                adaptins"
                /translation="AKYISQCIDEIKQELKQDNIAVKANAVCKLTYLQMLGYDISWAA
                FNIEVMSASKFTFKRIGYLAASQSFHEGTDVIMLTTNQIRKDLSSPSQYDTGVALTG
                LSCFVTPDLARDLANDIMTLMSHTKPYIRKKALIMYKVFLKYPESLRPAFPRLEKEI
                EDPDGVQSAAVNVICELARRNPKNYLSLAPLFFKLMTSSTNNWVLIKIIKLRLFGALTP
                LEPRLGKKLIEPLTNLIHSTSAMSLIYECVNTVIVAVLISLSSGMNHSASIOLCVOKL
                RILIEDSDONLKYLGLLAMSKILKTHPKSVQSHKDILLQCLDDKDDKDESIRLRALDLLYG
                MVSKKNLMEIVKKLMTHVDKAEGTTYRDELLTKIIDICSQSNYQYITNFEWISILVE
                LTRLEGTRHGHLIAAQMLDVAIRVRAIRKFAVSQMSALLLDSAHLLASSTQRNGICEVL
                YAAAWICGEFSEHLQEPHHTLEAMLRPRVTTLPGHIQAVYQNVVKLYASILQOKEQA
                GEAEGAQAVTQLMVDRLPQFVQSADLEVQERASCILQLVKHIQKLQAKDVPVAEEVSA
                LFAGELNPVAPKAQKVPVPEGLDLDAWINEPLSDSESEDERPRAVFEHEEQRRKHR
                PSEADEELARREARKQEQANNPFYIKSSPSPQKRYQDTPGVEHIPVVOIDLSVPLK
                VPGLPMSDQYVKLEEERRHRQKLEKDKRRKRRKRERKEKGKRRHSSLPTESDEDIAPAQ
                OVDIYTEEMPENALPSDEDKDPNDYRALDIDLOKPLADSEKLPIOKHNTETSKSP
                EKDVPMVEKKSKPKKKEKHKEKERDKEKKKEKEKRSPKPKKKHRKEKEERTGK
                KKSKCKQPPGSEEAAGEPVQNGAPEEEQLPPESSYSLLAENSYVKMTCDIRGSLOEDSQ
                VTVAIVLENRSSSTILKGMELSVLDSLNARMARPQGSSVHDGVPVPQLPPGVSNEAQY
                VFTIQSIVMAQKLKGTLSFIAKNDEGATHEKLDFRLHFSCSSYLITTPCYSDAFAKLL
                ESGDLSMSSIKVDGIRMSFQNLLAKICFHHFPSVVERVDSCASMYSRSIQOHHVCLLV
                KKGENSVSVDGKCSDSTLLSNLLEEMKATLAKC"
```

YLITTPCYSDAFAKLLESGDLSMSSIKVDGIRMSFQNLLAKICFHHFPSVVERVDSCA
SMYSRSIQOHHVCLLVKKGENSVSVDGKCSDSTLLSNLLEEMKATLAKC"

```
repeat_region   3435..3567
                /rpt_family="AluSq"
repeat_region   3570..3866
                /rpt_family="AluY"
repeat_region   3867..3906
                /rpt_family="Alu"
repeat_region   3908..4090
                /rpt_family="(TAAA)n"
repeat_region   4101..4266
                /rpt_family="AluSq"
repeat_region   complement(5793..6096)
                /rpt_family="FRAM"
repeat_region   complement(6348..6855)
                /rpt_family="AluJo"
repeat_region   6938..7227
                /rpt_family="MER9"
repeat_region   7255..7402
                /rpt_family="AluSg"
repeat_region   7403..7689
                /rpt_family="LIMB5"
repeat_region   7693..8049
                /rpt_family="AluSx"
repeat_region   8105..8399
                /rpt_family="LIMB5"
                /rpt_family="AluY"
```

```
RESULT 13
AF165142

Qy      653  atgaaaatgagccacacccttcca  675
             |||||||||||||||||||||||||
Db    10491  ATGAAAATGAGCCACACCTTCCA  10469
```

Query Match   0.9%;   Score 23;   DB 86;   Length 43514;
Best Local Similarity   100.0%;   Pred. No. 0.43;
Matches   23;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
LOCUS       AF165142      105490 bp    DNA             PRI       24-AUG-1999
DEFINITION  Homo sapiens chromosome 8 clone BAC 392c11 map 8p11-p12, complete
            sequence.
ACCESSION   AF165142
VERSION     AF165142.1  GI:5499748
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105490)
  AUTHORS   Schudy,A., Blechschmidt,K., Schillhabel,M., Baumgart,C., Menzel,U.,
            Weber,J., Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..105490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p11-p12"
                     /clone="BAC 392c11"
     exon            complement(84..140)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   complement(333..592)
                     /rpt_family="AluSx"
     repeat_region   716..761
                     /rpt_type=tandem
     repeat_region   complement(1299..1513)
                     /rpt_family="MIR"
     exon            3417..3519
                     /note="MZEF"
                     /evidence=not_experimental
     exon            complement(3617..4219)
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(4209..4252)
                     /rpt_family="Charlie1a"
     repeat_region   4302..4405
                     /rpt_family="Charlie1a"
     exon            complement(5085..5193)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(5315..5460)
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(5347..5460)
                     /rpt_family="FLAM_C"
     repeat_region   complement(5373..5537)
                     /rpt_family="AluJo"
     exon            complement(6488..6564)
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(6760..7029)
                     /rpt_family="AluY"
     repeat_region   7180..7227
                     /rpt_type=tandem
     repeat_region   complement(7232..7489)
                     /rpt_family="AluSx"
     exon            complement(7489..7563)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8092..8142
                     /note="Initial_exon"
                     /evidence=not_experimental
     exon            complement(8101..8149)
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            8306..8360
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   9096..9528
                     /note="GRAIL"
                     /evidence=not_experimental
```

```
     repeat_region   9191..9229
                     /rpt_family="Tigger2a"
     repeat_region   9223..9282
                     /rpt_type=tandem
     exon            9223..9282
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            10179..10635
                     /note="terminal_exon"
     repeat_region   10824..10879
                     /rpt_family="MER31B"
     repeat_region   10882..11145
                     /rpt_type=tandem
     repeat_region   11138..11339
                     /rpt_family="MER31-internal"
     repeat_region   11351..11994
                     /rpt_family="MER31-internal"
     repeat_region   11563..11606
                     /rpt_family="MER31-internal"
     repeat_region   12029..12223
                     /rpt_type=tandem
     repeat_region   12122..12742
                     /rpt_family="MER57-internal"
     repeat_region   12457..13937
                     /rpt_family="MER4-internal"
     repeat_region   12637..13415
                     /rpt_family="MER65-internal"
     repeat_region   12749..14506
                     /rpt_family="MER65-internal"
     exon            13011..13027
                     /rpt_family="MER31-internal"
     exon            14127..14173
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            14239..14331
                     /note="XPOUND"
                     /evidence=not_experimental
     exon            14439..14470
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   14513..14985
                     /note="XPOUND"
                     /evidence=not_experimental
     repeat_region   14986..15205
                     /rpt_family="MER31B"
     repeat_region   15583..15624
                     /rpt_family="MER31-internal"
     repeat_region   complement(15594..16226)
                     /rpt_type=tandem
     repeat_region   complement(15778..16081)
                     /rpt_family="AluYa8"
     repeat_region   complement(16414..16776)
                     /rpt_family="AluSc"
     repeat_region   complement(17761..18189)
                     /rpt_family="THE1B"
     repeat_region   complement(18043..18271)
                     /rpt_family="L1PA13"
     repeat_region   18286..18554
                     /rpt_family="L1"
     exon            18291..18397
                     /rpt_family="L1PA10"
     exon            18291..18397
                     /note="MZEF"
                     /evidence=not_experimental
     exon            18291..18397
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   complement(18661..19071)
                     /rpt_family="MLT1F"
     repeat_region   complement(19781..20183)
                     /rpt_family="MER83"
     exon            20013..20114
                     /note="GRAIL"
                     /evidence=not_experimental
```

```
repeat_region    complement(21158..21279)
                 /rpt_family="MIR"
repeat_region    21802..21906
                 /rpt_family="MER5A"
repeat_region    complement(22020..22325)
                 /rpt_family="AluJb"
repeat_region    22911..23014
                 /rpt_family="MLT1J"
repeat_region    23103..23196
                 /rpt_family="L2a"
exon             complement(23581..23725)
                 /note="GRAIL"
exon             23583..23663
                 /note="GRAIL"
                 /evidence=not_experimental
exon             24026..24185
                 /note="GRAIL"
                 /evidence=not_experimental
exon             complement(24093..24257)
                 /note="Terminal_exon"
exon             24114..24185
                 /note="XPOUND"
                 /evidence=not_experimental
exon             24201..24344
                 /note="GRAIL"
                 /evidence=not_experimental
repeat_region    complement(24884..24945)
                 /rpt_family="L1MA6"
repeat_region    complement(25524..25621)
exon             /note="XPOUND"
                 /evidence=not_experimental
repeat_region    25642..25699
                 /rpt_family="L2"
repeat_region    25837..25924
                 /rpt_type=tandem
repeat_region    26272..27029
                 /rpt_family="L1PA7"
exon             26733..26881
                 /note="GRAIL"
                 /evidence=not_experimental
exon             27656..27783
                 /note="GRAIL"
repeat_region    complement(27765..28027)
                 /rpt_type=tandem
repeat_region    28150..28450
                 /rpt_family="LTR37B"
repeat_region    28452..29399
                 /rpt_family="MER33"
repeat_region    28897..29188
                 /rpt_family="L1MA5"
repeat_region    29402..29593
                 /rpt_family="AluSx"
                 /rpt_family="Charlie5"
```

Query Match   0.98;   Score 23;   DB 89;   Length 105490;
Best Local Similarity   100.0%;   Pred. No. 0.44;
Matches 23;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy  1428 aagaagtcagtaccagaaatca 1450
         |||||||||||||||||||||||
Db 39697 AAGAAGTCAGTACCAGAAATCA 39719
```

RESULT 14

First section

COMMENT
```
          On Dec 22, 2000 this sequence version replaced gi:11878435.
          requests: clonerequest@sanger.ac.uk
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          Submitted (21-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL   Direct Submission
TITLE     Direct Submission
AUTHORS   Howden,P.
REFERENCE  1  (bases 1 to 64151)
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM  Homo sapiens
SOURCE    human.
KEYWORDS  HTG.
VERSION   AL139399.9  GI:11989998
ACCESSION AL139399
          Xq21.33-22.3, complete sequence.
DEFINITION Human DNA sequence from clone RP11-574A21 on chromosome
LOCUS     AL139399    64151 bp    DNA    PRI    21-DEC-2000
RESULT 15
```

```
Qy  1428 aagaagtcagtacccagaaatca 1450
         |||||||||||||||||||||||
Db 161565 AAGAAGTCAGTACCCAGAAATCA 161587
```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Best Local Similarity 100.0%; Pred. No. 0.45;
Query Match 0.9%; Score 23; DB 65; Length 190705;

```
ORIGIN
BASE COUNT    55947 a  36657 c  35874 g  60716 t  1511 others

         misc_feature   152329 .190705
                        /note="assembly_name:Contig22"
         misc_feature   15329
                        /note="assembly_name:Contig21"
         misc_feature   124653 .152228
                        /note="assembly_name:Contig21"
         misc_feature   102659 .124552
                        /note="assembly_name:Contig20"
         misc_feature   82660 .102558
                        /note="assembly_name:Contig19"
         misc_feature   67351 .82559
                        /note="assembly_name:Contig18"
         misc_feature   57164 .67250
                        /note="assembly_name:Contig17"
         misc_feature   48596 .57063
                        /note="assembly_name:Contig16"
         misc_feature   40814 .48495
                        /note="assembly_name:Contig15"
         misc_feature   30744 .40713
                        /note="assembly_name:Contig14"
         misc_feature   24816 .30643
                        /note="assembly_name:Contig13"
         misc_feature   19610 .24715
                        /note="assembly_name:Contig12"
         misc_feature   13422 .19509
                        /note="assembly_name:Contig11"
         misc_feature   8750 .13321
                        /note="assembly_name:Contig10"
         misc_feature   3906 .8649
                        /note="assembly_name:Contig9"
         misc_feature
                        /note="assembly_name:Contig8
                        clone_end:SP6
                        vector_side:left"
                        clone="RP11-104D16"
                        /chromosome="8"
                        /db_xref="taxon:9606"
                        /organism="Homo sapiens"
         source         1 .2064
                        1 .190705
                        Location/Qualifiers
FEATURES

     * 152329 190705: contig of 38377 bp in length.
```

Second section

```
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP11-574A21 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-574A21 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-274M8 is at 64052 in this sequence.
The true right end of clone RP1-267M20 is at 100 in this sequence.
```

```
FEATURES        Location/Qualifiers
     source     1 .64151
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="q21.33-22.3"
                /clone="RP11-574A21"
                /clone_lib="RPCI-11.2"
                1 .64151

ORIGIN
BASE COUNT    20073 a  11574 c  12196 g  20308 t
```

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Best Local Similarity 100.0%; Pred. No. 6.3;
Query Match 0.8%; Score 21; DB 90; Length 64151;

```
Qy  1351 ctgcagcttcccaacttccag 1371
         |||||||||||||||||||||
Db 24576 CTGCAGCTTCCCAACTTCCAG 24556
```

```
RESULT 16
LOCUS     AC009929    125403 bp    DNA    PRI    18-APR-2000
DEFINITION Homo sapiens chromosome CTD-2291F22, complete sequence.
ACCESSION AC009929
VERSION   AC009929.3  GI:6598820
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 125403)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE     Direct Submission
JOURNAL   Unpublished
REFERENCE  2  (bases 1 to 125403)
AUTHORS   DOE Joint Genome Center.
TITLE     Direct Submission
JOURNAL   Direct Submission
          Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  3  (bases 1 to 125403)
AUTHORS   DOE Joint Genome Center and Stanford Human Genome Center.
TITLE     Direct Submission
```

RESULT 17

```
Query Match
Best Local Similarity  100.0%;  Score 21;  DB 87;  Length 125403;
Matches 21;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy       745 ttaaagagggaggaaacttcttg 765
             ||||||||||||||||||||||
Db     78704 TTAAAGAGGAGGAAACTTCTTG 78724
```

```
LOCUS       AC027347   148807 bp  DNA    HTG    18-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2562I21, WORKING DRAFT
            SEQUENCE, 3 ordered pieces.
ACCESSION   AC027347
VERSION     AC027347.3  GI:9256715
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148807)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 148807)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7712058.
            -----
            Genome Center
            Center: Joint Genome Institute
            Center code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 814002
            Center clone name: CITB-E1_2562I21
            -----
```

```
JOURNAL     Submitted (17-DEC-1999) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 125403)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (01-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 125403)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   6  (bases 1 to 125403)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:6165130.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1..125403
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="CTD-2291F22"
BASE COUNT    39991 a  24654 c  23616 g  37142 t
ORIGIN
```

```
Qy       745 ttaaagagggaggaaacttcttg 765
             ||||||||||||||||||||||
Db     78704 TTAAAGAGGAGGAAACTTCTTG 78724
```

RESULT 18

```
Summary Statistics
Consensus quality:   147486 bases at least Q40
Consensus quality:   148415 bases at least Q30
Consensus quality:   148542 bases at least Q20
Estimated Insert size: 234950; agarose-fp estimation
Estimated insert size: 148757; sum-of-contigs estimation
Quality coverage: 7.48 in Q20 bases; agarose-fp estimation
Quality coverage: 11.81 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*     1      85201: contig of 85201 bp in length
* 85202      85301: gap of unknown length
* 85302     103160: contig of 17859 bp in length
* 103161    103260: gap of unknown length
* 103261    148807: contig of 45547 bp in length
FEATURES             Location/Qualifiers
     source          1..148807
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2562I21"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    46028 a  28099 c  28931 g  45549 t  200 others
ORIGIN
```

```
Query Match
Best Local Similarity  100.0%;  Score 21;  DB 70;  Length 148807;
Matches 21;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy       745 ttaaagagggaggaaacttcttg 765
             ||||||||||||||||||||||
Db      5835 TTAAAGAGGAGGAAACTTCTTG 5815
```

```
LOCUS       AL442223   162344 bp  DNA    HTG    24-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-426P2, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL442223
VERSION     AL442223.6  GI:11493327
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162344)
AUTHORS     Pavitt,R.
TITLE       Direct Submission
JOURNAL     Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11493090.
            -----
            Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -----
            Project Information
            Center project name: bA426P2
            -----
            Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 160707 bases at least Q40
```

RESULT 19

LOCUS   AP000719   196424 bp   DNA   PRI   22-FEB-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-757C15,
            complete sequences.
ACCESSION   AP000719
VERSION   AP000719.4 GI:13094220
KEYWORDS   HTG.
SOURCE   Homo sapiens DNA, clone:RP11-757C15.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Homo sapiens genomic DNA
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2 (bases 1 to 196424)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Qy   338 tgcactggtctcaatgtgggg 358
           |||||||||||||||||||||
Db 31695 TGCACTGGTCTCAATGTGGGG 31675

BASE COUNT   49214 a   32490 c   32393 g   48046 t   201 others
ORIGIN

Query Match   Best Local Similarity 100.0%; Score 21; DB 81; Length 162344;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES   Location/Qualifiers
     source   1..162344
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone_lib="RPCI-11.2"
                     /clone="RP11-426P2"
     misc_feature   1..33529
                     /note="assembly_fragment:02253
                     fragment_chain:1"
     misc_feature   33630..68844
                     /note="assembly_fragment:01727
                     fragment_chain:1"
     misc_feature   68945..162344
                     /note="assembly_fragment:01094
                     vector_side:right"
                     clone_end:T7

     1   33529: contig of 33529 bp in length
     33530 33629: gap of   100 bp in length
     33630 68844: contig of 35215 bp in length
     68845 68944: gap of   100 bp in length
     68945 162344: contig of 93400 bp in length.

*   NOTE: This is a 'working draft' sequence. It currently
*   consists of 3 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.

Insert size: 162144; sum-of-contigs
Insert size: 144616; agarose-fp
Quality coverage: 5.90x in Q20 bases; agarose-fp
coverage: 6.66x in Q20 bases; sum-of-contigs Quality

Consensus quality: 161514 bases at least Q30
Consensus quality: 161929 bases at least Q20

COMMENT
FEATURES   Location/Qualifiers
     source   1..196424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-757C15"
ORIGIN
On Feb 21, 2001 this sequence version replaced gi:9927288.

Qy   976 tttcagaggtaccaaggaaag 996
           |||||||||||||||||||||
Db 160132 TTTCAGAGGTACCAAGGAAAG 160152

BASE COUNT   54557 a   42826 c   43446 g   55595 t
ORIGIN

Query Match   Best Local Similarity 100.0%; Score 21; DB 91; Length 196424;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 20

LOCUS   AP002490   196613 bp   DNA   HTG   10-NOV-2000
AP002490/c
DEFINITION Homo sapiens chromosome 11 clone CMB9-17C1 map 11q12, WORKING DRAFT
            SEQUENCE, 25 unordered pieces.
ACCESSION   AP002490
VERSION   AP002490.2 GI:11138084
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens DNA, clone:CMB9-17C1.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 196613)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Homo sapiens 196,613 genomic DNA of 11q12
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2 (bases 1 to 196613)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-JUN-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
On Nov 11, 2000 this sequence version replaced gi:8547575.
COMMENT   -------------------- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            -------------------- Project Information
            Center project name: HumDraft11
            Center clone name: CMB9-17C1
            -------------------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 190034 bases at least Q40
            Consensus quality: 192516 bases at least Q30
            Consensus quality: 193548 bases at least Q20
            Insert size: 194213; sum-of-contigs
            Quality coverage: 8.73x in Q20 bases; sum-of-contigs
            --------------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            25 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)

are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

```
          1   27851: contig of  27851 bp in length
      27952   58349: contig of  30398 bp in length
      58450   76029: contig of  17580 bp in length
      76130   92048: contig of  15919 bp in length
      92149  103650: contig of  11502 bp in length
     103751  114318: contig of  10568 bp in length
     114419  123627: contig of   9209 bp in length
     123728  131291: contig of   7564 bp in length
     131392  139816: contig of   8425 bp in length
     139917  146953: contig of   7037 bp in length
     147054  153039: contig of   5986 bp in length
     153140  158083: contig of   4944 bp in length
     158184  163158: contig of   4975 bp in length
     163259  169022: contig of   5764 bp in length
     169123  171598: contig of   2476 bp in length
     171699  175793: contig of   4095 bp in length
     175894  178616: contig of   2723 bp in length
     178717  181798: contig of   3082 bp in length
     181899  184417: contig of   2519 bp in length
     184518  187490: contig of   2973 bp in length
     187591  189268: contig of   1678 bp in length
     189369  191568: contig of   2200 bp in length
     191669  193333: contig of   1665 bp in length
     193434  195156: contig of   1723 bp in length
     195257  196613: contig of   1357 bp in length.
```

NOTE: This is a 'working draft' sequence. It currently
consists of 25 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```
*          1   27851: contig of  27851 bp in length
*      27852   27951: gap of      100 bp
*      27952   58349: contig of  30398 bp in length
*      58350   58449: gap of      100 bp
*      58450   76029: contig of  17580 bp in length
*      76030   76129: gap of      100 bp
*      76130   92048: contig of  15919 bp in length
*      92049   92148: gap of      100 bp
*      92149  103650: contig of  11502 bp in length
*     103651  103750: gap of      100 bp
*     103751  114318: contig of  10568 bp in length
*     114319  114418: gap of      100 bp
*     114419  123627: contig of   9209 bp in length
*     123628  123727: gap of      100 bp
*     123728  131291: contig of   7564 bp in length
*     131292  131391: gap of      100 bp
*     131392  139816: contig of   8425 bp in length
*     139817  139916: gap of      100 bp
*     139917  146953: contig of   7037 bp in length
*     146954  147053: gap of      100 bp
*     147054  153039: contig of   5986 bp in length
*     153040  153139: gap of      100 bp
*     153140  158083: contig of   4944 bp in length
*     158084  158183: gap of      100 bp
*     158184  163158: contig of   4975 bp in length
*     163159  163258: gap of      100 bp
*     163259  169022: contig of   5764 bp in length
*     169023  169122: gap of      100 bp
*     169123  171598: contig of   2476 bp in length
*     171599  171698: gap of      100 bp
*     171699  175793: contig of   4095 bp in length
*     175794  175893: gap of      100 bp
*     175894  178616: contig of   2723 bp in length
*     178617  178716: gap of      100 bp
*     178717  181798: contig of   3082 bp in length
*     181799  181898: gap of      100 bp
*     181899  184417: contig of   2519 bp in length
*     184418  184517: gap of      100 bp
*     184518  187490: contig of   2973 bp in length
*     187491  187590: gap of      100 bp
*     187591  189268: contig of   1678 bp in length
*     189269  189368: gap of      100 bp
*     189369  191568: contig of   2200 bp in length
*     191569  191668: gap of      100 bp
*     191669  193333: contig of   1665 bp in length
*     193334  193433: gap of      100 bp
*     193434  195156: contig of   1723 bp in length
*     195157  195256: gap of      100 bp
*     195257  196613: contig of   1357 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..196613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q12"
                     /clone="CMB9-17C1"
     misc_feature    1..27851
                     /note="assembly_fragment"
     misc_feature    27952..58349
                     /note="assembly_fragment"
     misc_feature    58450..76029
                     /note="assembly_fragment"
     misc_feature    76130..92048
                     /note="assembly_fragment"
     misc_feature    92149..103650
                     /note="assembly_fragment"
     misc_feature    103751..114318
                     /note="assembly_fragment"
     misc_feature    114419..123627
                     /note="assembly_fragment"
     misc_feature    123728..131291
                     /note="assembly_fragment"
     misc_feature    131392..139816
                     /note="assembly_fragment"
     misc_feature    139917..146953
                     /note="assembly_fragment"
     misc_feature    147054..153039
                     /note="assembly_fragment"
     misc_feature    153140..158083
                     /note="assembly_fragment"
     misc_feature    158184..163158
                     /note="assembly_fragment"
     misc_feature    163259..169022
                     /note="assembly_fragment"
     misc_feature    169123..171598
                     /note="assembly_fragment"
     misc_feature    171699..175793
                     /note="assembly_fragment"
     misc_feature    175894..178616
                     /note="assembly_fragment"
     misc_feature    178717..181798
                     /note="assembly_fragment"
     misc_feature    181899..184417
                     /note="assembly_fragment"
     misc_feature    184518..187490
                     /note="assembly_fragment"
     misc_feature    187591..189268
                     /note="assembly_fragment"
     misc_feature    189369..191568
                     /note="assembly_fragment"
     misc_feature    191669..193333
                     /note="assembly_fragment"
     misc_feature    193434..195156
                     /note="assembly_fragment"
     misc_feature    195257..196613
                     /note="assembly_fragment"
BASE COUNT     55160 a     43875 c     44156 g     51021 t     2401 others
ORIGIN
```

```
Query Match
Best Local Similarity    0.8%;
Matches   21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

            Score 21;  DB 83;  Length 196613;
            Pred. No. 6.5;

Qy    976 tttcaagaggtaccaaggaaag 996
          |||||||||||||||||||||
Db  37392 TTTCAGAGGTACCAAGGAAAG 37372

RESULT  21
LOCUS       AC034179   220146 bp    DNA             HTG      12-NOV-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-797O11, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC034179
VERSION     AC034179.3  GI:11139352
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 220146)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 220146)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (04-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 12, 2000 this sequence version replaced gi:7524008.
            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----- Project Information -----
            Center project name: H_NH0797O11
            --------- Summary Statistics ---------
            Sequencing vector: M13; 95%
            Sequencing vector: plasmid; 5%
            Chemistry: Dye-primer ET; 95% of reads
            Chemistry: Dye-terminator Big Dye; 5% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 210550 bases at least Q40
            Consensus quality: 213061 bases at least Q30
            Consensus quality: 214675 bases at least Q20
            Insert size: 205000; agarose-fp
            Insert size: 218246; sum-of-contigs
            Quality coverage: 4.85 in Q20 bases; agarose-fp
            Quality coverage: 4.64 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6923: contig of 6923 bp in length
            *     6924     7023: gap of unknown length
            *     7024    10977: contig of 3954 bp in length
            *    10978    11077: gap of unknown length
            *    11078    17882: contig of 6805 bp in length
            *    17883    17982: gap of unknown length
            *    17983    22351: contig of 4369 bp in length
            *    22352    22451: gap of unknown length
            *    22452    28667: contig of 6216 bp in length
            *    28668    28767: gap of unknown length
            *    28768    35913: contig of 7146 bp in length
            *    35914    36013: gap of unknown length
            *    36014    43927: contig of 7914 bp in length
            *    43928    44027: gap of unknown length
            *    44028    51182: contig of 7155 bp in length
            *    51183    51282: gap of unknown length
            *    51283    63340: contig of 12058 bp in length
            *    63341    63440: gap of unknown length
            *    63441    85661: contig of 22221 bp in length
            *    85662    85761: gap of unknown length
            *    85762   105434: contig of 19673 bp in length
            *   105435   105534: gap of unknown length
            *   105535   134835: contig of 29301 bp in length
            *   134836   134935: gap of unknown length
            *   134936   198011: contig of 63076 bp in length
            *   198012   198111: gap of unknown length
            *   198112   200255: contig of 2144 bp in length
            *   200256   200355: gap of unknown length
            *   200356   203493: contig of 3138 bp in length
            *   203494   203593: gap of unknown length
            *   203594   205664: contig of 2071 bp in length
            *   205665   205764: gap of unknown length
            *   205765   208649: contig of 2885 bp in length
            *   208650   208749: gap of unknown length
            *   208750   212298: contig of 3549 bp in length
            *   212299   212398: gap of unknown length
            *   212399   216138: contig of 3740 bp in length
            *   216139   216238: gap of unknown length
            *   216239   220146: contig of 3908 bp in length.
FEATURES             Location/Qualifiers
     source          1..220146
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-797O11"
     misc_feature    1..6923
                     /note="assembly_name:Contig10"
     misc_feature    7024..10977
                     /note="assembly_name:Contig11"
     misc_feature    11078..17882
                     /note="assembly_name:Contig12"
     misc_feature    17983..22351
                     /note="assembly_name:Contig13"
     misc_feature    22452..28667
                     /note="assembly_name:Contig14"
     misc_feature    28768..35913
                     /note="assembly_name:Contig15"
     misc_feature    36014..43927
                     /note="assembly_name:Contig16"
     misc_feature    44028..51182
                     /note="assembly_name:Contig17"
     misc_feature    51283..63340
                     /note="assembly_name:Contig18"
     misc_feature    63441..85661
                     /note="assembly_name:Contig19"
     misc_feature    85762..105434
                     /note="assembly_name:Contig20"
     misc_feature    105535..134835
                     /note="assembly_name:Contig21
                     vector_side:right"
     misc_feature    134936..198011
                     /note="assembly_name:Contig22"
     misc_feature    198112..200255
                     /note="assembly_name:Contig3"
     misc_feature    200356..203493
                     /note="assembly_name:Contig4"
     misc_feature    203594..205664
                     /note="assembly_name:Contig5"
     misc_feature    205765..208649
                     /note="assembly_name:Contig6"
     misc_feature    208750..212298
                     /note="assembly_name:Contig7"
     misc_feature    212399..216138
                     /note="assembly_name:Contig8"
```

```
        misc_feature    216239..220146
                        /note="assembly_name:Contig99"
BASE COUNT       66033 a    41274 c    41175 g    69151 t    1913 others
ORIGIN

Query Match       0.8%;    Score 21;    DB 71;    Length 220146;
Best Local Similarity    100.0%;    Pred. No. 6.6;
Matches 21;    Conservative 0;    Mismatches 0;    Indels 0;    Gaps 0;

Qy    745 ttaaagaggagaaacttcttg 765
          |||||||||||||||||||||
Db  51116 TTAAAGAGGAGAAACTTCTTG 51136
```

```
RESULT 22
LOCUS       AB046207    1176 bp    mRNA    VRT    11-APR-2001
DEFINITION  Anguilla japonica mRNA for 28kDa-1d apolipoprotein, complete cds.
ACCESSION   AB046207
VERSION     AB046207.1  GI:13591607
KEYWORDS
SOURCE      Anguilla japonica liver cDNA to mRNA.
  ORGANISM  Anguilla japonica
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Anguilliformes;
            Anguilloidei; Anguillidae; Anguilla.
REFERENCE   1 (sites)
  AUTHORS   Kondo,H.; Kawazoe,I.; Nakaya,M.; Kikuchi,K.; Aida,K. and Watabe,S.
  TITLE     The novel sequences of major plasma apolipoproteins in the eel
            Anguilla japonica(1)
  JOURNAL   Biochim. Biophys. Acta 1531 (1-2), 132-142 (2001)
  MEDLINE   21175592
REFERENCE   2 (bases 1 to 1176)
  AUTHORS   Watabe,S., Kondo,H., Kawazoe,I., Nakaya,M., Kikuchi,K. and Aida,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2000) Shugo Watabe, The University of Tokyo,
            Graduate School of Agricultural and Life Sciences; Yayoi 1-1-1,
            Bunkyo, Tokyo 113-8657, Japan
            (E-mail:awatabe@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-7520,
            Fax:81-3-5841-8166)
FEATURES             Location/Qualifiers
     source          1..1176
                     /organism="Anguilla japonica"
                     /db_xref="taxon:7937"
                     /tissue_type="liver"
     CDS             18..797
                     /codon_start=1
                     /product="28kDa-1d apolipoprotein"
                     /protein_id="BAB40964.1"
                     /db_xref="GI:13591608"
                     /translation="MKFIVIALVSFTGCQQANVLRSNEPTPQLELVKDAFWDYFVKAS
                     QTAQGALKIFGESELAQQVNAKIKESVEVVQQYRRIVQEQVIVISDELHKKLSEHAEQ
                     LSESLQPDINEVRVQLEPLAEKLSANIQQOMQKVRPVLDPYTESLDIRALERALRRMH
                     RRLFMSVEQLLSQLLEHLQDQLGPSTEELKRKVEESVQEYREFVLPLENILGNEILNK
                     LFMFPRTVYPIVDGLREKLDPYIQGLESQLTALWESSANSA"
BASE COUNT       333 a    262 c    286 g    295 t
ORIGIN

Query Match       0.8%;    Score 20;    DB 8;    Length 1176;
Best Local Similarity    100.0%;    Pred. No. 22;
Matches 20;    Conservative 0;    Mismatches 0;    Indels 0;    Gaps 0;

Qy   1373 agagcgtgcaggagtacagg 1392
          ||||||||||||||||||||
Db    601 AGAGCGTGCAGGAGTACAGG 620
```

```
RESULT 23
LOCUS       ATU23796    1573 bp    mRNA    PLN    21-JUL-1995
DEFINITION  Arabidopsis thaliana IAA-amino acid hydrolase homolog ILL2 (ILL2)
            mRNA, complete cds.
ACCESSION   U23796
VERSION     U23796.1  GI:902790
KEYWORDS
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1573)
  AUTHORS   Bartel,B. and Fink,G.R.
  TITLE     ILR1, an amidohydrolase that releases active indole-3-acetic acid
            from conjugates
  JOURNAL   Science 268 (5218), 1745-1748 (1995)
  MEDLINE   95312867
REFERENCE   2 (bases 1 to 1573)
  AUTHORS   Bartel,B. and Fink,G.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1995) Bonnie Bartel, Whitehead Institute, 9
            Cambridge Center, Cambridge, MA 02142, USA
COMMENT     On Jul 21, 1995 this sequence version replaced gi:887788.
FEATURES             Location/Qualifiers
     source          1..1573
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /chromosome="5"
     CDS             1..1320
                     /gene="ILL2"
                     /note="ILR1-like gene; IAA-amino acid hydrolase homolog"
                     /codon_start=1
                     /product="ILL2"
                     /protein_id="AAC49016.1"
                     /db_xref="GI:902791"
                     /translation="MALNKLLSLTFQLLLFLLSVSSESPWIAEDTSOIQTKLLEFAKS
                     PEVFDWMKIRRKIHENPELGYEELETSKLIRSELELIGIKYRPYVAITGYVIGYICTG
                     EPPEVALRADMDALPIQEGVEWEHKSKIPGKRMHACHGDGHVTMLGAAKILHEHRHL
                     QGTVVLIFQPAEEGLSGGAKKMREEGALKNVEAIFGIHLSARIFPGKRAASRAGSFLAGA
                     GVFEAVTTGKGGHAAIPHHTIDPVVAASSIVLSLCQQLVSREFDPLDSKVVTVSKVNGG
                     NAFNVIPDSTIGGTLRAFTGFTQLQORVKEVITKOAAVHRCNASVNLTPNGREPMPP
                     TVNNKDLYKQFRKVVRDLLGQEAFVERAPVMGSEDFSYFATTIPGHFSLLGMQDEFNG
                     YASSHSPLYRINEDVLPYGAAIHASMAVQYLKEKASKGSVSGFHEEL"
     gene            75..1320
                     /gene="ILL2"
BASE COUNT       452 a    332 c    363 g    426 t
ORIGIN

Query Match       0.8%;    Score 20;    DB 14;    Length 1573;
Best Local Similarity    100.0%;    Pred. No. 22;
Matches 20;    Conservative 0;    Mismatches 0;    Indels 0;    Gaps 0;

Qy   1703 ctttgggaaagccgcttcac 1722
          ||||||||||||||||||||
Db    609 CTTTGGGAAAGCCGCTTCAC 628
```

```
RESULT 24
LOCUS       AF047031    4918 bp    DNA    PLN    03-MAY-1999
DEFINITION  Arabidopsis thaliana IAA-amino acid hydrolase (ILL1 and ILL2)
            genes, complete cds.
ACCESSION   AF047031
VERSION     AF047031.1  GI:2921828
KEYWORDS
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 4918)
  AUTHORS   Bartel,B. and Fink,G.R.
  TITLE     ILR1, an amidohydrolase that releases active indole-3-acetic acid
```

```
                                    from conjugates
JOURNAL   Science 268 (5218), 1745-1748 (1995)
MEDLINE   95312867
REFERENCE 2  (bases 1 to 4918)
  AUTHORS  Davies,R.T., Goetz,D.H., Lasswell,J., Anderson,M.N. and Bartel,B.
  TITLE    IAR3 encodes an auxin conjugate hydrolase from Arabidopsis
  JOURNAL  Plant Cell 11 (3), 365-376 (1999)
  MEDLINE  99172052
REFERENCE 3  (bases 1 to 4918)
  AUTHORS  Bartel,B.
  TITLE    Direct Submission
  JOURNAL  Submitted (08-FEB-1998) Biochemistry and Cell Biology, Rice
           University, 6100 S. Main St., Houston, TX 77005, USA
FEATURES             Location/Qualifiers
     source          1..4918
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /chromosome="5"
                     /db_xref="taxon:3702"
                     /map="between ngal29 and LFY3"
     mRNA            join(<143..496,605..730,805..1119,1201..1323,1408..>1806)
                     /gene="ILL1"
                     /product="IAA-amino acid hydrolase"
     gene            <143..>1806
                     /gene="ILL1"
                     /note="ILR1-like gene 1"
     CDS             join(143..496,605..730,805..1119,1201..1323,1408..1806)
                     /gene="ILL1"
                     /codon_start=1
                     /function="auxin conjugate hydrolase"
                     /product="IAA-amino acid hydrolase"
                     /protein_id="AAC04865.1"
                     /db_xref="GI:2921829"
                     /translation="MAINNFLTFQLLLLLRVSSESPWIVAGDVSRIPINFLELAKSP
                     EVFDSMYRIRRKIHENPELGYEEFETSKFIRSELDLIGVKYRPPVAITGIGYIGTGE
                     PPFVALRADMDALPIQEAVEWEHKSKNPGKMHACGHDGHVAMLLGAAKILQQHRQHLQ
                     GTVVLIFQPAEEGLSGAKMMREEGALKNVEAIFGIHLSPRTPFGKAASLAGSPMAGAG
                     AFEAVITGKGGHAAIPQHTIDPVVAASSIVLSLQHLVSRETDPSDSKVVTVTKVNGGN
                     AFNVIPDSITIGGTLRAFTGTTQLQRRIKEIITKQAAVHRCNASVNLAPNQNQPMPPT
                     VNNMDLYKFKKVVRDLLGQEAFVEAVPEMGSEDFSYFAETIPGHFSLLGMQDETQGY
                     ASSHSPHYRINEDVLPYGAAIHATMAVQYLKDKASKGSSVSGFHDEL"
     gene            <2596..>4329
                     /gene="ILL2"
                     /note="ILR1-like gene 2"
     mRNA            join(<2596..2952,3121..3246,3323..3637,3718..3840,
                     3931..>4329)
                     /gene="ILL2"
                     /product="IAA-amino acid hydrolase"
     CDS             join(2596..2952,3121..3246,3323..3637,3718..3840,
                     3931..4329)
                     /gene="ILL2"
                     /codon_start=1
                     /function="auxin conjugate hydrolase"
                     /product="IAA-amino acid hydrolase"
                     /protein_id="AAC04866.1"
                     /db_xref="GI:2921830"
                     /translation="MAINKLLSLTFQLLLFLLSVSSESPWIAEDTSQIQTKLLEFAKS
                     PEVFDMWVKIRRKIHENPELGYEELETSKLIRYRPVAITGIVGIVGTG
                     EPPFVALRADMDALPIQEGVEWEHKSKIAGKMHACGHDGHVTMLLGAAKILHEHRHL
                     QGTVVLIFQPAEEGLSGAKKMREEGALKNVEAIFGIHLSARIPFGKAASRAGSFLAGA
                     GVFEAVTTGKGGHAAIPQHTIDPVVAASSIVLSLQQLVSREDPLDSKVVTVSKVNGG
                     NAFNVIPDSITIGGTLRAFTGTTQLQQRVKEVITKQAAVHRCNASVNLTPNGREPMPP
                     TVNNKDLYKQFKKVVRDLLGQEAFVERAPVMGSEDFSYFAETIPGHFSLLGMQDETNG
                     YASSHSPLYRINEDVLPYGAAIHASMAVQYLKEKASKGSVSGFHEEL"
     variation       3082..3083
                     /gene="ILL2"
                     /note="6 base insertion in ecotype Landsberg erecta"
                     /replace="tgaatt"
BASE COUNT     1422 a    991 c    997 g   1508 t
ORIGIN
```

Query Match   0.8%;   Score 20;   DB 12;   Length 4918;

Best Local Similarity   100.0%;   Pred. No. 23;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy      1703 ctttgggaaagccgcttcac 1722
             ||||||||||||||||||||
Db      3448 CTTTGGGAAAGCCGCTTCAC 3467

RESULT   25
LOCUS       AB014516      6718 bp   mRNA            PRI       06-FEB-1999
DEFINITION  Homo sapiens mRNA for KIAA0616 protein, partial cds.
ACCESSION   AB014516
VERSION     AB014516.1  GI:3327045
KEYWORDS    .
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
            SK plus clone:HG03623.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6718)
  AUTHORS   Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
            Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
REFERENCE   2  (sites)
  AUTHORS   Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
            Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. X.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (3), 169-176 (1998)
  MEDLINE   98403880
FEATURES             Location/Qualifiers
     source          1..6718
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HG03623"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..1906
                     /gene="KIAA0616"
     CDS             <1..1906
                     /gene="KIAA0616"
                     /codon_start=2
                     /product="KIAA0616 protein"
                     /protein_id="BAA31591.1"
                     /db_xref="GI:3327046"
                     /translation="HNQKQAEEFAAFEEVMKDLSLTRAARLQKSQYLQLQPSRGQY
                     YGGSLPNVNQIGGCTMDLPFQPSGFLGEALAAAPVSLIPFQSGGLDTSRTTRHHGLVD
                     RVRERGRLGSPHRPLSVDKHGRQADSCPYGTMYLSPPADTSWRRTNSDSALHQSTM
                     TPTOPESPSSSGSODVHQKRVLLIVPGMEETTSEADKNLSKQAWDTKTGSRPKSCEV
                     STGNLAANLTHLCIGGACGMSTPGSSPOHRPAGVSPLSLSTEARRQGASPTLSPLSP
                     ITQAVAMDALSLEQQLPYAFFTQAGSQQPPDQPPQQPPPPPPPASOQPPOPPPQOAPVRL
                     PPGPLLPSASLTRGPQOPPPLAVTVVPSSLPQSPPENPGQPSMGIDIASAPALOQYRTS
                     AGSPANQSPTSPVSNQGFSPGSSPQHTSGTLGSVPGDAYEOQMARQANALSHOLEQF
                     NMMENAISSSLYSPGSTLNYSQAAMMGLTGSHGSLPDSOQLGYASHSGIPNILTVT
                     GESPPSLSKELTSSLAGVGDVSFDSDSOQPPLDELKIDPLTLDGLHMLNDPDMVLADPA
                     TEDTFRMDRL"
BASE COUNT     1208 a   2237 c   1930 g   1343 t
ORIGIN
```

Query Match   0.8%;   Score 20;   DB 85;   Length 6718;
Best Local Similarity   100.0%;   Pred. No. 23;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
RESULT 26
Qy      1601 gggtccttgggcaccctggct 1620
             ||||||||||||||||||||||
Db      3731 GGGTCCTGGGCACCCTGGCT 3750

SC7B7/c
LOCUS       SC7B7          13800 bp    DNA        BCT       02-DEC-1997
DEFINITION  Streptomyces coelicolor cosmid 7B7.
ACCESSION   AL009199
VERSION     AL009199.1  GI:2661675
KEYWORDS    ABC-transport system; hydrolase; periplasmic solute-binding
            protein; transcriptional repressor protein; transmembrane transport
            protein.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 13800)
  AUTHORS   Murphy,L. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13800)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-1997) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrel@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE   3  (bases 1 to 13800)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research. 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl.  CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 7B7 is in
            the AseI-B fragment, between 7H1 and 1C3.
FEATURES             Location/Qualifiers
     source          1..13800
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
```

```
     gene            complement(1..1878)
                     /clone="cosmid 7B7"
                     /gene="SC7B7.01c"
     CDS             complement(1..1878)
                     /gene="SC7B7.01c"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical ATP/GTP binding protein"
                     /protein_id="CAA15781.1"
                     /db_xref="GI:2661676"
                     /db_xref="SPTREMBL:O50498"
                     /translation="MQYRVRHRWGRASVPWGPAVVMTAFALAVGSQGAAVALPGAPAKADREFA
                     SSFEADDPVPDWTDTVDTDRDGGKRASGVDGGYSTGIPGNVTENVTGVRASAENTGGR
                     EVKENLVDNEPATKWLTFAPTGWVEEFDLDAPTKIAKYALTSANDHERDPLDWTLKGS
                     ADGENWQTLDTRSGEHFDGRRQTRTYELAETAEFQHFRLEITKNNGAGDALQLADVQL
                     ATSDTEVPTPEDMLSFVDEGPVGSFPTAKANAGFGRHALRYAGRHFHADGRAYSYNKVF
                     DVNVKVDRRSELSYKIPPSMADGDLDYDATNVSVDLAFTDGTYLSDLRATDQHGFPLT
                     PRGQGDDAKILYVNQNNSVKSGIGSVAAGKTVDRILLAYDSPKGPAKFRGWEDDVAIER
                     AAPERPKAHLSDYALITTRGTNSSGSFSRGNNFPPATAVPHGFGNFWFTPVTNAGSLSWIYE
                     YASGNNDDNLPTVQAFSASHEPSPWMGDRQOTFQVMPSAASGTPDTGPDARELAFRHEN
                     ETARPYYYGVRFENGVKAEMTPTDHAAMMRFTYPGDDASVIFDNVNDQAGLITDKENG
                     TFTGYSDVRSGGSAGATRLFVYGEFDSRVTGGDSSGVKGHLRFDAGRDHSVTMRIATS
                     LI"
     misc_feature    complement(787..810)
                     /gene="SC7B7.01c"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     RBS             complement(1890..1894)
                     /gene="SC7B7.01c"
                     /note="possible RBS upstream of SC7B7.01c"
     RBS             2428..2431
                     /note="possible RBS upstream of SC7B7.02"
     gene            2440..3849
                     /gene="SC7B7.02"
     CDS             2440..3849
                     /gene="SC7B7.02"
                     /note="SC7B7.02, unknown, len: 469; has similarity to
                     signal sequence of Mycobacterium leprae hypothetical
                     protein TR:E334742 (EMBL:Z98604) MLCB2052.27 (445 aa),
                     fasta scores; Opt: 75 z-score: 199.2 E(): 0.0062, 23.3%
                     identity in 335 aa overlap. Contains N-terminal signal
                     sequence and appropriately positioned PS00013 Prokaryotic
                     membrane lipoprotein lipid attachment site"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC7B7.02"
                     /protein_id="CAA15782.1"
                     /db_xref="GI:2661677"
                     /db_xref="SPTREMBL:O50499"
                     /translation="MTIRAGSLDRRTLLRGAIATAAMGSPAVACSSPSSEDKESDSGP
                     KGEKSANNPFGAAAANSTVEAAIFDGGYGTDYDYANQVLGSQVKGLKVQKPVVDIAP
                     OLQPREVGGNPPDLIDNSGEDQIGFIGILDQLEELDLFFASTYEGKKIADIVFPGVK
                     DPGTKDKFVALNYVMTVYGVWYSKTLFEENGWTPPKTWDEALDLGQEBAKKGKTLFV
                     HGKEAATYYIRTLLIDSAIKEGGDEVRIALENLEKGCWSHPAVGQVIKVMETMVKQRMF
                     VPGGSGTQFQKRAQAIVSNDQKALLYPSGGWIENEMKKATKADFQMTGIPSMTLTDKPA
                     LPYEALRAAAGEPFIVPKQGKNPAGGKEVLRAMLSEKAAANFSKTHKLAPTIVKGTVPA
                     DGYGSTALVSQTKMLEAAGTNIFNYMFVETYGLNTDQLVPWNSFLAGDLDGKGLTSAL
                     QKISDKVREDDSVDKVKVS"
     misc_feature    2497..2529
                     /gene="SC7B7.02"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
                     attachment site"
     gene            3857..4918
                     /gene="SC7B7.03"
     CDS             3857..4918
                     /gene="SC7B7.03"
                     /note="SC7B7.03, probable integral membrane transport
                     protein, len: 353 aa: similar to many members of the MalFG
                     subfamily of the binding protein-dependant transport
                     system membrane proteins eg. MSMF_STRMU Q00750 multiple
                     sugar-binding transport system permease protein (290 aa),
```

```
LOCUS       AC003107    46275 bp    DNA       PRI      18-NOV-1997
DEFINITION  Human DNA from chromosome 19-specific cosmid R30064 containing the
            COMP gene, genomic sequence, complete sequence.
ACCESSION   AC003107
VERSION     AC003107.1  GI:2623749
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46275)
  AUTHORS   Lamerdin,J.E.; McCready,P.M., Adamson,A.W., Burkhart-Schultz,K.,
            Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J.,
            Danganan,L., Christensen,M., Bruce,R., Quan,G., Montgomery,M.,
            Ow,D., Kobayashi,A., Olsen,A.O. and Carrano,A.V.
  TITLE     Sequence analysis of an ~1 Mb region containing the MEF2B gene in
            19p12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46275)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Cosmid R30064 overlaps cosmid F19807 to the left and cosmid R32469
            to the right.
FEATURES             Location/Qualifiers
     source          1..46275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R30064"
                     /chromosome="19"
                     /map="19p12 between UBA52 and D19S451"
                     /cell_line="5HL2-B"
                     /clone_lib="LL19NCO3 cosmid library"
                     /note="LL19NCO3 R chromosome 19 cosmid library
                     flow-sorted chromosomes from hybrid 5HL2-B, which carries
                     chromosome 19 as its only human chromosome."
     misc_feature    817..1037
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 93.000"
     repeat_region   complement(3847..3899)
                     /rpt_family="MER7"
     repeat_region   complement(3910..4198)
                     /rpt_family="Alu"
     repeat_region   complement(4214..4315)
                     /rpt_family="MER7"
     misc_feature    6213..6337
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 0, quality: excellent, score: 93.000"
     repeat_region   complement(6488..6785)
                     /rpt_family="Alu"
     repeat_region   complement(6934..7242)
                     /rpt_family="Alu"
     repeat_region   complement(7269..7545)
                     /rpt_family="Alu"
     repeat_region   complement(7989..8304)
                     /rpt_family="Alu"
     misc_feature    9294..9409
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 65.000"
     misc_feature    9543..9602
                     /note="DDS similarity to AA378771 EST91505 Synovial
                     sarcoma Homo sapiens cDNA 5' end (1..60); 100% identity."
     misc_feature    9676..10030
                     /note="DDS similarity to AA209234 zq85f01.r1 Stratagene
                     hNT neuron (#937233) Homo sapiens cDNA clone 648409 5'
                     (1..351); Score: 660 Identity: 346/351 (98%)."
                     complement(10540..11045)
                     /note="DDS similarity to AA211938 zq85f01.s1 Stratagene
                     hNT neuron (#937233) Homo sapiens cDNA clone 648409 3'
                     (1..351); Score: 951 Identity: 497/501 (99%)."
```

```
     misc_feature
                     Map and sequence oriented from telomere to centromere.
```

```
     misc_feature    4505..4591
                     /gene="SC7b7.03"
                     /note="PS00402 Binding-protein-dependent transport systems
                     inner membrane component signature"
     RBS             4904..4908
                     /gene="SC7b7.03"
                     /note="possible RBS upstream of SC7b7.04"
     gene            4915..5823
                     /gene="SC7b7.04"
     CDS             4915..5823
                     /gene="SC7b7.04"
                     /note="SC7b7.04, probable integral membrane transport
                     protein, len: 301 aa; similar to many members of the MALFG
                     subfamily of the binding-protein-dependant transport
                     systems membrane proteins eg. TR:O06111 (EMBL:X660092)
                     putative transport system permease protein from
                     Clostridium perfringens (275 aa), fasta scores; opt: 590
                     z-score: 796.5 E(): 0, 33.6% identity in 265 aa overlap.
                     Contains PS00402 Binding-protein-dependent transport
                     systems inner membrane comp signature"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transmembrane transport protein"
                     /protein_id="CAA15784.1"
                     /db_xref="SPTREMBL:O50501"
                     /translation="MSVIAADRKPASDRKPVSDRKLARAAASSDRRFAAISHALLILW
                     SVIVIVPMLWVLMSSFKSTGEILSSPFSLPDHWRFENYANAWTDANIGKYFLNSVIVV
                     VSALIIVMLLGAMCAVVLAREFPGRRLIYIVMLAGLTFWLPLFFQLQNFGLL
                     NTRPGLILTYVAFALEPTMFFLYSFRSLPHDVYEAALIDGAGDWRAFFQVWLPMARP
                     GMAAVAIFNFLGLWNQFLLPVALNTDQDKWVLTQGMAAYASSQVYDIDYGALFAAIVV
                     TVVPVLLVYCVFQRRIAGSVSQGTFR"
     misc_feature    5470..5556
                     /gene="SC7b7.04"
                     /note="PS00402 Binding-protein-dependent transport systems
                     inner membrane comp signature"
     stem_loop       5851..5900
                     /note="hairpin loop with 23bp stem"
     RBS             6044..6047
                     /note="possible RBS upstream of SC7B7.05"
     gene            6058..7269
                     /gene="SC7B7.05"
     CDS             6058..7269
                     /gene="SC7B7.05"
```

```
fasta scores: opt: 393 z-score: 636.8 E(): 2.6e-28, 32.5%
identity in 280 aa overlap. Contains PS00402 Binding-protein-dependent transport systems inner membrane comp signature"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transmembrane transport protein"
                     /protein_id="CAA15783.1"
                     /db_xref="SPTREMBL:O50500"
                     /translation="MKDTIPTAETASRRPEPAARGRPRRRKLTFDRVTEFLAFLGVP
                     LAIFVIFVLPFGQAIFWGMTDWRGFSPDYNFVGFPDNFTKMFQDDIPIKALRNVALLA
                     AFVPLVTLTLALGVAVAITLGGPSKGPVRGIRGASFYRIISFPPYVVPAIIVGLIWAQ
                     MYDPNAGLLNGVLTGLGLDQDFTPAWLGEKAAAMPAVMFVIVWGLVGFYAVLFIAAIK
                     GVPGELYEAAKIDGAGRFRTIISITLPAIRDSVQTAYIYLGIAALDAFVYVQAMVPNG
                     GPDNSTLTISQRLFNVAFAPAKQQFGYATAMGVVLAAVTLVFPAALVFLVNRLTGGGEGES
                     KRKAPGSRARRAAAKGGAR"
```

```
Qy        77  ccgcccgccggaagcggccg  96
              |||||||||||||||||||
Db      3925  ccgcccgcccgccgagcgccg  3906
```

```
RESULT  27
Db      AC003107
Qy        77  ccgcccgccggaagcggccg  96
              |||||||||||||||||||
Db      3925  ccgcccgcccgccgagcgccg  3906

Query Match     0.88;   Score 20;  DB 3;  Length 13800;
Best Local Similarity  100.00%;  Pred. No. 23;
Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

misc_feature　11916..12187
/note="DDS similarity to T82171 yd95g06.r1 Homo sapiens
cDNA clone 116026 5' similar to contains TAR1 repetitive
element.Score: 478 Identity: 262/274 (95%)."

Other overlapping matches:
(11993..12236) predicted exon, program:
grail2exons_human_1.3, frame: 2, quality: excellent,
score: 81.000
(12176..11917) DDS similarity to T95404 ye43g01.r1 Homo
sapiens cDNA clone 120528 5'; Score: 471 Identity:
258/263 (98%).-(12713..11917) DDS similarity to T84000
yd66d10.r1 Homo sapiens cDNA clone 113203 5'; Score: 400
Identity: 257/275 (93%).-(11965..12181) DDS similarity to
T9520 Ye43g01.s1 Homo sapiens cDNA clone 120528 3'
similar to gb:M64241 QM PROTEIN (HUMAN).Score: 334
Identity: 203/218 (93%)."

misc_feature　12251..12355
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 52.000-DDS similarity to
AA378771 EST91505 Synovial sarcoma Homo sapiens cDNA 5'
end (61..165). 96% identity.
"

misc_feature　13676..13705
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 54.000"

misc_feature　14355..14457
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 77.000"

misc_feature　15674..15795
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 65.000"

misc_feature　15709..15764
/note="DDS similarity to AA378771 EST91505 Synovial
sarcoma Homo sapiens cDNA 5' end (166..221); 99%
identity.

misc_feature　16450..16630
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"

misc_feature　17229..17316
/note="DDS similarity to T95918 ye42a04.r1 Homo sapiens
cDNA clone 120366 5'. Score: 163 Identity: 86/87 (98%)."

misc_feature　18023..18187
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"

misc_feature　18533..18757
/note="DDS similarity to AA252605 zs14b02.s1 NCI_CGAP_GCB1
Homo sapiens cDNA clone 685131 3' end. Score: 424
Identity: 225/225 (100%)."

misc_feature　20682..22267
/note="similarity to overlapping ESTS:
(20682..21075) AA350958 EST58512 Infant brain Homo
sapiens cDNA 5' end; Score: 776 Identity: 391/394 (99%).
(20863..21259) R87396 ym88h07.r1 Homo sapiens cDNA clone
120366 5'. end; Score: 748 Identity: 391/401 (97%).
(20874..21249) AA349930 EST57030 Infant brain Homo
sapiens cDNA 5' end; Score: 748 Identity: 375/376 (99%).
(21125..21574) R19558 yg26h07.r1 Homo sapiens cDNA clone
33726 5'. Score: 811 Identity: 440/457 (96%).
(21589..22030) H30788 yo79f06.r1 Homo sapiens cDNA clone
184163 5' similar to gb:M33326 NONSPECIFIC CROSS-REACTING
ANTIGEN NCA-95 (HUMAN); contains Alu repetitive element.
(22007..22267) AA223326 EST24948 Cerebellum II Homo
sapiens cDNA 5' end; Score: 479 Identity: 252/262 (96%)."

repeat_region　complement(21512..21626)
/rpt_family="Alu"

misc_feature　22506..22980
/note="DDS similarity to R88263 ym90h09.s1 Homo sapiens
cDNA clone 162241 3'. Score: 837 Identity: 455/478
(95%).-Other overlapping matches:

misc_feature
(23013..22497) H05926 yl71f07.s1 Homo sapiens cDNA clone
43353 3'; Score: 872 Identity: 497/527 (94%).
(22971..22534) W95528 AA446916 zw85a01.s1 Soares total fetus
Nb2H8 9w Homo sapiens cDNA clone 783720 3'; Score: 857
Identity: 436/437 (99%)
(22971..22561)W95490 ze02g07.s1 Soares fetal heart NbHH19W
Homo sapiens cDNA clone 357852 3';
(22566..22949) W95528 ze02g07.r1 Soares fetal heart
NbHH19W Homo sapiens cDNA clone 357852 5';Score: 674
Identity: 366/382 (95%).
(22996..22634) T33668 EST58674 Homo sapiens cDNA 3' end
similar to None. .; Score: 710 Identity: 359/363 (98%)."
complement(23576..23770)
/note="DDS similarity to N94385 zb76d06.s1 Soares
senescent fibroblasts NbHSF Homo sapiens cDNA clone 305515
3'

gene
similar to gb:Z19585 THROMBOSPONDIN 4 PRECURSOR (HUMAN);
(194..1); 92% identity.-Other overlapping matches:
(23724..23770) predicted exon, program:
grail2exons_human_1.3, frame: 1, quality: good, score:
74.000-(23770..23664) DDS similarity to W36954 mb82e10.r1
Soares mouse p3NMF19.5 Mus musculus cDNA clone 335946 5'
similar to
gb:Z19585 THROMBOSPONDIN 4 PRECURSOR (HUMAN); gb:M86620
Mouse thrombospondin 3 (MOUSE) (307..423) 63% identity."
complement(23683..32077)
/gene="COMP"
/note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR

CDS
(COMP)"
complement(join(23724..23770..23683..24002,25000..25172,
25705..25901,26307..26355,26482..26660,26774..26955,
27048..27100,27341..27459,28229..28458,29020..29127,
29218..29322,29400..29558,29647..29721,29968..30105,
30750..30922,31370..31421,31658..31743,31999..32077))
/gene="COMP"
/note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR
(COMP)"
/codon_start=1
/product="COMP_HUMAN"
/protein_id="AAB86501.1"
/db_xref="GI:2623750"
/translation="MVPDTACVLLLTLAALGASGQGCGQSPLGSDLGPQMLRELQEFNAA
LQDVRELLRQQVREITFLKNTVMEDCAGMQQSVRIGLPSVRPLLHCAPGFCYPGYSG
IQTESGARCGPCPAGFTGNGSHCTDVNECNAHPCFPRVRCINTSPGFRCEACPPGYSG
PTHQGVGLAFAKANKQVCTDINECETGQHNCVPNSVCINTRGSFQCGPCQPGFVGDQA
SGCQRRAQRFCPDGSPSECHEHADCVLERDGSRSCVCAVGWAGNGILCGRDTDLDGFP
DEKLRCPERQCRKDNCVTVPNSGQEDVDRDGIGDACDPDAPGDGVPNEKDNCPLVRNP
DQRNTDEDKWGDACDNCRCSQKNDDKDTDDPGRGDACDDDIDGDRIRNQADNCPRVPN
SDQKDSDGDGIGDACDNCPQKSNPDQADVDHDFVGDACDSDQDQDGDGHQSRDNCPT
VPNSAQEDSDHDGQGDACDDDDNDGVPDSRDNCRLVPNPGQEDARDGVGDVCQDDF
DADKVVDKIDVCPENAEVTLTDFRAFQTVVLDPEGDAQIDPNWVVLNQGREIVQTMNS
DPGLAVGYTAFNGVDFEGTFHVNTVTDDDYAGFIFGYQDSSSFYVVMWKQMEQTVWQA
NPFRAVAEPGIQLKAVKSSTGPGEQLRNALMHTGDTESQVRLLWKDPRNVGWKDKKSY

Qy　1601 gggtcctgggcaccctggct 1620
|||||||||||||||||||
Db　20016 GGGTCCTGGGCACCCTGGCT 20035

Query Match　100.0%; Score 20; DB 85; Length 46275;
Best Local Similarity 0.8%;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
AUTHORS     Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
            and Tabata,S.
TITLE       Structural analysis of Arabidopsis thaliana chromosome 5. III.
            Sequence features of the regions of 1,191,918 bp covered by
            seventeen physically assigned P1 clones.
JOURNAL     DNA Res. 4 (6), 401-414 (1997)
MEDLINE     98162728
REFERENCE   2  (bases 1 to 79590)
AUTHORS     Nakamura,Y.
TITLE       Direct Submission
JOURNAL     Submitted (06-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
            Address for correspondence: kaos@kazusa.or.jp.
COMMENT     For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MXC9
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers. Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            splicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
FEATURES             Location/Qualifiers
     source          1..79590
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MXC9"
                     /clone_lib=Mitsui P1"
     CDS             complement(3423..3875)
                     /note="contains similarity to S1 self-incompatibility
                     protein
                     gene_id:MXC9.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10024.1"
                     /db_xref="GI:9759373"
                     /translation="MKNILKTQVHVVVIYLLIKIAFSQVKPDFDVAWSTSKMVRITNR
                     LGDGLVLNLHCKSADDLGLKILAPNGSWSKFRTSIVGTLFYCHFTWPGQSKWFNIY
                     YDDRDGVRSHISCINCIWDISIOGPCMFSESDHAFNICYDWNGNLRT"
     CDS             complement(5454..5903)
                     /note="gene_id:MXC9.2
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10025.1"
                     /db_xref="GI:9759374"
                     /translation="MLKTQVHVVVILLIQIAFSOVKPGGSDLKWSTLKSVVTITNRLG
                     DGSTLKLHCKSADDLGLKILAPNGSWSFKRPSIVPGVTLFFCHFTWPGQSKWFNIY
                     DDRDGVRMGIPCYCIWDIGKRYGPCRFSEIDDAFNICYDWNGNRRS"
     CDS             complement(join(6686..6926,7008..7210,7299..7592,
                     7870..7993,8085..9427))
                     /note="emb|CAB87679.1
                     gene_id:MXC9.3
                     strong similarity to unknown protein"
                     /codon_start=1
```

<!-- right column -->

```
                     /evidence=not_experimental
                     /protein_id="BAB10026.1"
                     /db_xref="GI:9759375"
                     /translation="MAEOKSSNGGGGGGGGGDVVINVPVEEASRRSKEMASPESEKGVPFS
                     KSPSPEISKLVGSPNKPPRAPNQNNVGLITQRKSFARSVTSKPKSRFVPDSCPVDTSIL
                     EEEVREQILGAGFSFRSRASPNNKSNRSVGSPAPVTPSKVVVEKDEDEEIYKVKLNREM
                     RSKISTLALIESAFFVVILSALVASLTINVLKHHTFWGLEVRMWCVLVMVIFSGMLVT
                     NWFMRLIVPLIETNFLLRRKVLYFVHGILKKSVQVFIWLCLILVAWILLFNHDVKRSPA
                     ATKVLKCITRTLIISILCGAFFWLVKTLLKILAANFNVNNFFDRIQDSVFHQVVLQTL
                     SGLPLMERAERVGREPSTGLHLSFATVVAVVRKGTVKEHVKHKMRKRVVSAWTMRV
                     LMEAVRTSGLSTISDTLDETAYGEGKEQOADREITSEMEALAAAXHVFRNVAQFFNYI
                     EEEDLLRRMIKEEVDLVFPLFDGAAEIGRITRKAFTEWVVKYVTSRRALAHSLNDTKT
                     AVKOLNKIVTAILMVVTVIWILLLLEVAATTKVLLFFSTQLVALAFIIGSTCKNLFESI
                     VFVFVMHPYDVGDRCVVDGVAMLVEEMNLLTTVFLKLNNEKVYYPNAVLATKRPISNYF
                     RSPNMGETVEFSISFSTPVSKIAHLKERIAEYLEQNPQHWAPVHSVVKEIENMKLK
                     MALYSDHTITFQENRERNLRRTELSLAIKRMLEDLHIOYTLLFQODINLTKKN"
     CDS             complement(13439..17695)
                     /note="gene_id:MXC9.4"
                     /codon_start=1
                     /pseudo
                     /evidence=not_experimental
                     /product="retroelement pol polyprotein-like"
                     join(18207..18234,18300..19381)
                     /note="contains similarity to NPK1-related protein kinase
                     gene_id:MXC9.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10027.1"
                     /db_xref="GI:9759376"
                     /translation="MIENDDSPQHPSLFFTDFTRANSHVRVLHHKFLGKGVIGSVDLI
                     RYTKTDGSSLQAAVKTSVAEDLLEYDALKREIQILSELKGPNIVICYDGLLEEDFNE
                     HGHKVYKLLLEYANEGSLSSFMENYPDRKLPDMIRDFTRMILEGLVSMHSHGYVHCD
                     LKSDNLLIFSRKDSASCELKIFDFGNCRQVGGEVPDHWKSDYPVVGTPEPESFFDGVAKKI
                     LDIWSLGCLVLKIYTGEQPWERVTSVDVFNFLSDGEAPNIPEVVPCDAREFIETCFAR
                     EHEKRGTASELLLHPFLCOKQPLIVKLKLRIKOASKITMGIREKPLTLKMFLSKPPOL
                     KKLLSKILRLKKMIHPKPLELDSSFVSYSSDLCLL"
     gene            19892..22342
                     /note="gbl|AAF19552.1
                     gene_id:MXC9.6
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10028.1"
                     /db_xref="GI:9759377"
                     /translation="MVTRLRLVSRSSRXATVKFTDSVSACSCRRLFSASTDEPESOP
                     EQAPPTNPVTDGEKLRNLRVLLQQONRIETARGVLSSLLRSDSTPFASPKELSAFSLS
                     SPSLKHDFSYLLLSVLLNESKMISEAADLFFALRNEGIYPSSDSLTLLLDHLVRTKQF
                     RVTINVFLNILESDFPRSKFMYGKAIQAAVKLSDVGKGLELENRMKHDRIYPSVFIYN
                     VLIDGLCKGKRMNDAEQLFDEMLARRLLPSLITYNTLIDGYCRAGNPEKFKVRERMK
                     ADHIEPSLITFNTLLKGLFKAGMWEDAENVLKDMXDLGFVPDAFTFSILFDGYSSNEK
                     AERALGYVETAVDSGVKMNAYTCSILLNALCEGKIEKAEEILGREMAKGLVPNEVIY
                     NTMIDGYCRKGDIVGARMKIEAMEKOGMKPDHLAYNCLIRRFCELGEMENAEKEVNKM
                     KLRGVSPSVETYNILIGYGRKYEFDKCFDILKEMEDNGTMPMNNVSVGTLINCLCKGS
                     KLLEAQIVKRDMEDRGVSPKVRIYNMLIDGCCSGKGKIEDAFRRSKEMLKKGIELNLVT
                     YNTLIDGLSMTGRLSEAEDLLLEISRKGIKPDVFTYNSLISGYGFAGNWQRCILALYEE
                     MKRSGIKPTLKTYHLLISLCTKEGIELTERLFGEMSLKPDLLVYNGVLHCYAVHGDME
                     KAFNLKQMIEKSIGLDKTYTNSLILGOLKVGKLCEVRSLIDEMNAREMEPADTYNI
                     IVKGHCEVKDYMSAYVWYREMQEKGFLLDVCIGNELVSGLKEEWRSKEAEIVISEMNG
                     RMLGDVTVPEDLSATEKL1"
                     join(22987..23072,23172..23288,23378..23453,23623..23752,
                     23874..23987,24073..24236)
                     /note="emb|CAB64729.1
                     gene_id:MXC9.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="elongation factor 1B alpha-subunit"
                     /protein_id="BAB10029.1"
                     /db_xref="GI:9759378"
                     /translation="MAVTFSDLHTERGLKTLEEHLAGKTYIISGDQLSVDDVKYYAAVL
                     ENPGDGFPNASKWDSVASHLAKSFPGDAGPHPEEPAADGDGD
                     DDDIDLFADETEDEKRAAEEREAAKKDTKKTKESGKSSVLLEVKPWDDETDMKKLEE
                     AVRSVOMPGLTWGASKLVPVGYGIKKLTIMMTIVDDLVSVDNLIEDHLTSEPNNEYIQ
                     SVDIVAFNKI"
```

RESULT 29

LOCUS　　　　AB013392　　　8429 bp　　　DNA　　　　　PLN　　　27-DEC-2000
DEFINITION　Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MIK19.
ACCESSION　　AB013392 BA000015
VERSION　　　AB013392.1　GI:3128139
KEYWORDS
SOURCE　　　　Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
　　　　　　　clone:MIK19.
ORGANISM　　Arabidopsis thaliana
　　　　　　　Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
　　　　　　　Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
　　　　　　　Brassicales; Brassicaceae; Arabidopsis.
REFERENCE　　1　(sites)
AUTHORS　　Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Miyajima,N.
　　　　　　　and Tabata,S.

COMMENT　　Address for correspondence: kaos@kazusa.or.jp
　　　　　　　For the latest information on annotation of this clone, please see
　　　　　　　http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MIK19
　　　　　　　Genes with similarity to proteins in the databases are described in
　　　　　　　'product' or 'note' qualifiers. Genes that have no significant
　　　　　　　protein similarity are described as 'unknown protein'.
　　　　　　　The software programs used to predict genes include: Grail
　　　　　　　(Informatics Group, Oak Ridge National Laboratory,
　　　　　　　http://compbio.ornl.gov/Grail-1.3/),
　　　　　　　http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
　　　　　　　GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
　　　　　　　NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
　　　　　　　Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
　　　　　　　SplicePredictor (Volker Brendel, Stanford University,
　　　　　　　http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
　　　　　　　Genes encoding tRNAs are predicted by tRNAscan-SE
　　　　　　　(Sean Eddy, Washington University School of Medicine, St. Louis,
　　　　　　　http://genome.wustl.edu/eddy/tRNAscan-SE/).
　　　　　　　This sequence may not be the entire insert of this clone. It may be
　　　　　　　shorter because we remove overlaps between neighboring submissions.
　　　　　　　The 5' clone is MKN22 and the 3' clone is MPI10.
FEATURES　　　　　　　　　　　Location/Qualifiers
　source　　　　　1..84129
　　　　　　　/organism="Arabidopsis thaliana"
　　　　　　　/strain="Columbia"
　　　　　　　/db_xref="taxon:3702"
　　　　　　　/chromosome="5"
　　　　　　　/clone="MIK19"
　　　　　　　/clone_lib="Mitsui P1"
　CDS　　　　join(941..1855,1946..2080,2340..2609)
　　　　　　　/note="gene_id:MIK19.1"
　　　　　　　pir||T06019
　　　　　　　similar to unknown protein"
　　　　　　　/codon_start=1
　　　　　　　/evidence=not_experimental
　　　　　　　/protein_id="BA09874.1"
　　　　　　　/db_xref="GI:10176760"
　　　　　　　/translation="MRKLFFFLCTHVPKQRAKFTRIMESCQKKQKLETPIELSDMPDD
LIFKIFSFLPFPKEDLATRFISEYGKGLWNPDNAIFDDDSLDMTCMKSFVGSLMSKD
AQILDSLHLKRKYYIASDINFLVQLGVNRSMRELRIDFFGKTLELPCCLSTCTILKV
LVLDHLNIMSVPGWFRLPSVETLQLSSVTGGSNYPSLIRLCSVHERLVVDQNRKDL
VNINVPTLRSLSIDNRRGHVPLIGSFWINVPSLMFLNIKIDFPQTIPFESMAPMIKANIE
VAHDQSQSVKFIESLTSTRHLCLCSPTSENPYPNGTPKFSNIVLHKLKLCTCSAGWQNLLA
CMLNDAPNLRSLTLKLRQKSDVNPRKVWEKPTVVPECLSTRLELIKWRDYEGTEHEKD
MVGYILANAAFLQRAFFSTKDRNCQDSRFSELQSMERVSEICEFVFD"
　CDS　　　　complement(join(3056..3148,3239..3284,3384..3607,
　　　　　　　3708..3890,3970..4197,4294..4428,4528..4612,4749..4825))
　　　　　　　/note="gene_id:MIK19.2"
　　　　　　　/codon_start=1
　　　　　　　/evidence=not_experimental
　　　　　　　/product="protein kinase MEK1 homolog"
　　　　　　　/protein_id="BA09875.1"
　　　　　　　/db_xref="GI:10176761"
　　　　　　　/translation="MYKIKSNLKQLKLSVPAQESPISSFLIASGFFHDGDFLLNQKGL
RLTSDEKQSRQSDSKELDFEITAEDLETVKVIGKGSGGVVOLVRHKWVGKFFAMKVIQ
MNIQEEIRKQIVQELKINQASSQCPHVVVCVHSFYHNGAFSLVEYMDRGSLADVIRQ
VKTILEPYLAVVCKQVLQGLVYLHNERHVIHRDIKPSNLLVNHKGEVKISDFGVSASL
ASSMGQRDTEVGTIVNYMSPERISGSYDYSSDIWSLGMSVLECAIGRFPYLESEDQON
PDSFYELLAAIVENPPTAPSDQFSPEFCSEFSVSACIQKDPPARASSLDLLSHPFIKKF
EDKDLDGILVGTLEPPVNILR"
　CDS　　　　join(5726..5780,5997..7280,7375..7408,7494..7641)
　　　　　　　/note="gene_id:MIK19.3"

Qy　1898　atttggaagagttcagacc　1917
　　　　|||||||||||||||||||||
Db　4335　ATTTGGAAGAGTTCAGACC　4354

Query Match　100.0%;　Score 20;　DB 12;　Length 79590;
Best Local Similarity　0.8%;　Pred. No. 24;
Matches 20;　Conservative 0;　Mismatches 0;　Indels 0;　Gaps 0;

　CDS　　　　complement(24733..26592)
　　　　　　　/note="gene_id:MXC9.8"
　　　　　　　pir||T02643
　　　　　　　strong similarity to unknown protein"
　　　　　　　/codon_start=1
　　　　　　　/evidence=not_experimental
　　　　　　　/protein_id="BAB10030.1"
　　　　　　　/db_xref="GI:9759379"
　　　　　　　/translation="MSSASKMKSRDKKVMNDSQKTPSKASGSMGGSSVVGGYNPLLG
FQTIESLSATGSSSLHHNGRFRHIDEDSGGDADCSVSNNGSWSGESGEHSKFSVYPST
SIAAKQEIPGADNKRDKMRLKNERKHQRQKEKRAQELHERCQYLMSRKLEVHAQL
IISMGIPHRATYALMLNEGKIEESINWLFDDGGAKVADKKLDPTSGNLKLDISQELG
RLELETKYKCTKQDVERAVVTAEGDIKEAEELLRRQKODQSTASRKPEDISSTSVN
NSKVPSVLTSQNTVGQLQPNSGMYPAGREEALDRKNLGYPRGSSYISGESENQSVNSI
MNGRTRMWTNGNGMGVAMGVQEAALVGNGREWTSPFFEGKDLFSLSRQYVPPSL"
ERIHMKLQOWMKLQQNAPLEENKRIPYQQTPLSRSTEETHYVTALGDQFKRLQODMRE
PVWMWQOQOQORSQSANTNVLPVSTMNASFTGAAAAGSGWHPANRSEAAOSNGYLP
SRSLPPNDLNSNLMYQOLQVOTQGQVNNNGHRMAGSSAPPAVAPAASLGFSGFGSL
AGSSSSGLDWNTDGSLGHLDYNNIDWSLDKGLACVRPSOQYVAASSPYSAASSPYEAH
join(28348..28450,28599..28744,28826..28902,28993..29080,
29161..29336,29392,29485..29566,29659..29708,
29854..29910,29996..30074,30190..30248,30406..30463,
30561..30658)
/protein_id="BAB10031.1"
/db_xref="GI:9759380"
/product="cystatin"
/note="gene_id:MXC9.9"
/codon_start=1
/evidence=not_experimental
/translation="MLMCFWVTILSAANRRLSHLPPIACSRGIDQEDEEKESRELLPH
KNDENATSRSSSSVDSGCLKDYQQEETYKTSFKTVALCVGTAVAFGIGILKEGVGK
ASEFFAGYILEQSLSVDNLFVFVLVFKYFKVPILMYQNKVLTYGIAGAIVFRFLILLG
TAILQKFFAVNLLAAVLLYSSFKLFASEEDDIGLSDNFIVKTICQRFIPVTSSYDGNR
FPTKHDGIIKATPLLTVAVIELSDIAFAVDSIPAVFGVFGVTRDFPIVLTSNLFAILGLR
SLYTLISEGMDELEYLQPSIAVVLGFGIGVKMILDFFGFHISTEASLGVVALSLSTGVL
LSLTNKSSDS"
/note="contains similarity to transmembrane transport
protein"
complement(join(31799..31984,32321..32440))
/note="gene_id:MXC9.10"
/codon_start=1
/evidence=not_experimental
/product="cystatin"
/protein_id="BAB10032.1"
/db_xref="GI:9759381"
/translation="MADQQAGTIVGGVRDIDANANDLQVESLARFAVDEHNKNENLTL
EYKRLLGAKTQVVAGTMHILTEVEVADGETNKVYEAKVLEKAWENLKQLESFNLHDV"

/emb|CAA03929.1

CDS
/codon_start=1
/evidence=not_experimental
/product="beta-1,3-glucanase-like protein"
/protein_id="BAB09876.1"
/db_xref="GI:10176762"
/translation="MARDFKLIFSISILLLLDCCYGGKGVGVCYGRSADDLPTPSKVV
QLIQQHNIKYVRIYDPNQVLKAFGNTSIELMIGVPNSDLNAFSQSOSNVDTVLKNSV
LPYYTPKITYITVGAERSTDDPHINASSFVVPAMQNVLTALEKVGLSRRIKVSTTLSL
GILSRPPSAGAFNSSYAYFLRPMLEFLAENKSPFMIDLYPYAYRPSPNNVSLDYV
LFESSEVIDPNTGLIYKNMPDAVDALIYALTALNFRTIKIMVTETGWPTRGSPKEK
AAASSDNAETYNSNIIRHVVTNQGTPAKPGEAMNVIFSLFNENRKAGLDSERNWGLF
YPDQTSVYQLDFTGKSNGFHSNSSGTNSSGSSNSWCIASSKASERDLKGALDWACGPG
NVDCTAIOPSQPCFQPDTIVSHASFVENSYFQQNRATDVACSFGGAGVKVKNKDPSYDK
CIYITAGGNKTRATNARALTSSASTPRGNELLQWILKLCLMITSLFFSLQTMNSQAL"
complement(join(8087..8221,8288..8425,8793..8915))

CDS
/note="gene_id:MIK19.4"
/codon_start=1
/evidence=not_experimental
/product="profilin-like protein"
/protein_id="BAB09877.1"
/db_xref="GI:10176763"
/translation="MSWQTYVDDHLMCDVAGNRLTAAAILGQDGSVWAQSNNFPQVKP
EEIQGIKDFTPGTLAPTGLFGGNKYMVIOGEPNAVIRGKKGAGGVTIKKTTLALV
FGIYDEPMTPGQCNMVVENLGEYLIESGL"
join(13604..13668,13748..14074,14151..14235)

CDS
/note="gene_id:MIK19.5"
/codon_start=1
/evidence=not_experimental
/product="retroelement pol polyprotein-like"
/protein_id="BAB09878.1"
/db_xref="GI:10176764"
/translation="MILYFIQKFVQEKIWQTYFQILSIVHHIOEDCGMYAGKKTPTIM
YEENAAYIAQLKDGYIKGDRKHILPKFFYTHDLQKSGDVRVLQIRLNDNLADLFTKA
LPTATFRKLVYRIGLRCLKLNLDEYTHRGEQHFKYFTSYGTIMVIRGSVMNKDGCPL"
join(14943..15139,15345..15429,15493..15582,15665..15758,
15841..15923,16024..16182)

CDS
/note="contains similarity to unknown protein
gb|AAD15447.1
gene_id:MIK19.6"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09879.1"
/db_xref="GI:10176765"
/translation="MTDETEEDDTTQORSSRNDGVSKNKGKGFKGDKAKRALIGAGGR
ILFYPLLIYNLVRFKLQSOFRCYAVDYRNFELFTKISEYOFGSDFESMQKDVPRLKKL
GVGGVITLNEPYETLVPSSLSAYEMEHLVIPIPRDYLFAPSIVDILAVNFINKNALL
GKTTVHCKAGRGRSTVVLCYLIEHKSMTVAAAFEHVRSIRPRVLLHPSQRKVSLDR
FIDMLHQRDNRITKMKT"
complement(join(16726..16824,17130..17372,17821..18332,
18890..19082,19231..19344))

CDS
/note="gene_id:MIK19.7
pir||T13434
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09880.1"
/db_xref="GI:10176766"
/translation="MKNSKCNLIDSKLEEHHHLCGSKHCPCGGRMIQAATKPNWVGLP
AGVKFDPTDQELIEHLEAKVKGKEENKWSSSHPLIDEFIPIDGEDGICYTHPQKLP
GVTRDGLSKHFHPKSRAYTIGTRKRKIIQTDHDSELTGSSETRWHKTGKTRPVMIN
GQRGCKKLVLYLTYTNFGKNRRPEKTNMWHQYHLGINEEEREGELVVSKIFYQTOPRQ
CVSNTMWSDHHGSKDVIGIGVGDEISSVAATLQSLSGSGDVVSRVNMHPHTRSFDEGTA
EASKGRENQHVSGTCEEVHDGIITSMSSHHMIHDHHNQHHQIGDRREFHMSSSYPMT
PTITSOHESIFHVTSTMPFQGSVVWFGIRPNYGLHSYVRRKLFLTNLCI"
join(22516..22617,22808..22885,22935..23115..23324,
23408..23566,23709,23800..23998,24082..24344,
24431..24431,24742..24933)

CDS
/note="gene_id:MIK19.8"
/codon_start=1
/evidence=not_experimental
/product="pyrophosphate-dependent
phosphofructo-1-kinase-like protein"
/protein_id="BAB09881.1"

CDS
/db_xref="GI:10176767"
/translation="MSSPRSNKPKIVNGPGGYILQDVPHLIDYLPDLPTYPNPLQDNP
AYSVKQYFVHADDSVPEKVVHHDGPRGVHFRRAGPROKVYFESDEVHACIVTCGGL
CPGLNTVIREVVSSLSYMYGVKRLLGIGYVSLGGYRGFYAKNTIPLNSKVVNDIHKRGG
TIIGTSRGGHDTNKIVDSTQDRGINQVYIIGGDGTQRGASVIFEEIRRRLKVAVVGI
PKITDNDIPVIDKSFGFGDTIAVEEAQRAINAAHVEAESNENGIGFEVKLMGRYSGYIAMY
ATLASRDYDCCLLPESPFYLEGEGLFIEIERRLKDHGHMVIVLAEGAGQDLLMCKSME
STPMDASGNKLLKDVGLWLSQSIKDHFKKNKMWNLKIYIDPTYMIRAVPSNASDNVYC
TLLAQSAVHGAMAGYTGYTSGLVNGRQTYIPFYFRITETQNNVVITDRWARLLSSTNQ
PSFLGPKDTSEEKKELPETPLLDDGAVIIPPVVRKEVTK"
complement(join(25739..25765,25861..25942,26048..26126,
26210..26279,26372..26494..26652,26704..26807..26866,
26966..27170,27390..27474,27561..27690,27813..27843))

CDS
/note="gene_id:MIK19.9
pir||T06010
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/product="IAA-amino acid hydrolase homolog 1 precursor"
/protein_id="BAB09882.1"
/db_xref="GI:10176768"
/translation="MNISVENPVVHEDSTTQKTGELRLDSDIPMSKISSDDEVFLAP
EMNAFGROPFRDYTDTNSEROKSVEHFYATCDTHNQTLDEYGVKMRSEYGKLDKMWNIWE
CCELSKEVVDESDPDLDEPQIOHLLQSARAIRKDYPNEDWLHLTALIHDLGKVLTLPQ
FGGLPQWAVVGDTFPVGCAFDESNVHHKYFMENEDPNNPKYNTKAGIYSEGCGLENVL
MSWGHDDYMYLVAKENGSTLPSPGLFITRYHSFYPLHKAGAYTHLMNEEDKENLKWLH
VENKYDLYSKSKVHVNVEKVRPKYYMSLIKKYFPENLRW"
join(29030..29383,29492..29617,29692..30006,30088..30210,
30295..30693)
sp|P54969*
/note="gene_id:MIK19.10
/codon_start=1
/evidence=not_experimental
/translation="MALNNFLTFQLLLLLRVSSESPWIVAGDVSRIPINFLELAKSP
EVFDSMVRIRRKIHENPELGYEFETSKFIRSELDLIGVKYRFPVAITGIGIGTGE
PPFVALRADMDALPIQEAVEWEHKSKNPGKMHACGHDGHVAMLLGAAKILQQHROHLQ
GTVVLIFQPAEEGLSGAKMREEGALKNVEAIFGIHLSPRTPGCKAASLAGSFMAGAG
AFEAVITGKGGHAAIPQHTIDPVVAASSIVLSLQHLVSRETDPSDSKVVTVTKVNGGN"

Qy   1703 cttgggaaagcccgcttcac 1722
            ||||||||||||||||||||||
Db   32335 CTTTGGGAAAGCCCGCTTCAC 32354

Query Match   0.8%;   Score 20;   DB 12;   Length 84129;
Best Local Similarity 100.0%;   Pred. No. 24;
Matches   20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
COMMENT      lemcke@mips.biochem.mpg.de.mayer@mips.biochem.mpg.de Project
             Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
             Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
             E-mail: michael.bevan@bbsrc.ac.uk
             Information on performance of analysis and a more detailed
             annotation of this entry and other sequences of chromosomes 3, 4
             and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1..103638
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
     misc_feature    1..8764
                     /note="overlap with BAC T22P22. Please refer to this BAC
                     for annotation and analysis"
     exon            complement(8151..8375)
                     /gene="F14F18_10"
                     /number=1
     gene            complement(8151..8375)
                     /gene="F14F18_10"
     gene            8151..9346
                     /gene="F14F18_10"
     CDS             complement(join(8151..8375,8495..8551,8650..8678,
                     8890..9346))
                     /gene="F14F18_10"
                     /note="similarity to predicted proteins in Synechococcus ,
                     Cyanophora and other organisms"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB87657.1"
                     /db_xref="GI:7573351"
                     /translation="MAATVLGHFSPITSSLKPPSRHRRFLHHNLSPKTPLQSSLIKFG
                     AENSEPQPRPLPETDCEVPPEEQQPINEYQSLSTSFPSWASGDLIEYSTRLFLTGAS
                     FAFFVGLPSVSWFGSSIGPEYEPVKRILAASSSGIFVVTLAVVRMYLGWAYVDEETGWYD
                     GQVWVKTPEVLARDRLLGSFSFSVKPVLARLKNTLVILGLSLILVINLGDSPIATSYRTY
                     RDPRDRSSLPIPGAYNDETARTFEPEAFCGEPSSDLL"
     intron          complement(8376..8494)
                     /gene="F14F18_10"
                     /number=1
     exon            complement(8495..8551)
                     /gene="F14F18_10"
                     /number=2
     intron          complement(8552..8649)
                     /gene="F14F18_10"
                     /number=2
     exon            complement(8650..8678)
                     /gene="F14F18_10"
                     /number=3
     intron          complement(8679..8889)
                     /gene="F14F18_10"
                     /number=3
     exon            complement(8890..9346)
                     /gene="F14F18_10"
                     /number=4
     gene            complement(join(11457..11637,11711..11760,11866..12033,
                     12133..12211,12407..12507,12606..12674,12823..12863,
                     13089..13356,13438..14056,14134..14237,14416..14856,
                     14953..15048,15122..15565))
                     /gene="F14F18_20"
     exon            complement(11457..11637)
                     /gene="F14F18_20"
                     /number=1
     gene            11457..15565
                     /gene="F14F18_20"
     CDS             complement(join(11457..11637,11711..11760,11866..12033,
                     12133..12211,12407..12606..12674,12823..12863,
                     13089..13356,13438..14056,14134..14237,14416..14856,
                     14953..15048,15122..15565))
                     /gene="F14F18_20"
                     /note="the mRNA covering this gene is only partial. This
                     has allowed the definition of the carboxy terminus of the
                     protein. The amino terminus remains predicted but
                     prediction is supported by different gene models."
                     /codon_start=1
```

```
                     /product="MAP3K delta-1 protein kinase"
                     /protein_id="CAB87658.1"
                     /db_xref="GI:7573352"
                     /translation="MSKMKHLLRKLHLHIGGSSGVGGGFADHHRLLDDSTRPMIDPSPILS
                     TSPSPASTSSVSSSGFGNASTTMPRLDTFEPVGRDLTAVDGVDFNLMEEEYQVQLAMA
                     ISVSDPDPRENADTAQLDAAKRISLGVSAPVTDADSAVDFLSLRTWSGCSISEGLKEL
                     GSWLPAVFSSLRVFLNPNCGHKVINYDQKVRDGFYDVGITTSNLSLSQGRMPLLVDLQ
                     AISISDNVDVEVILVNRLIDPELQELERVFALASECPDFAPGQVSSSDLITQKIANIVV
                     EQMGGPVENADEALRRWMLRSYELRNSLNTTILPLGRVNVGLARHRALLPKVLADRIN
                     LPCMLVKGSYTGTGDGAVNLIKLDDKRSPNSFFCSEYIIDLMGAPGALIPSEVPSSF
                     SDRCAVEKHQTERFEHDFGKLMHSQOISGENMPPFSGKPTCAQKVKVKNVSKYVISAA
                     KNPEPAQKLHAVLLESGASPPPDLFMDINPHNLRGKNLLQELRDESSNSMWSGIPCYP
                     EKVAEQLRESERNPTAESYQQSVEVDLSMKRNFDLDNTGRASSSSENMEVGTADGSAV
                     CDSHDQGINPLLGEAAKWEIMWEDLQIGERIGIGSYGEVTRAEWNGTEVAVKKFLDQD
                     FSGDALTQFRKSEIEIMLRLRHPNVVLFMGAVTRPPNFSILTEFLPRGSLYRLLHRPNH
                     QLDEKRRMRALDVAKGMNYLHTSHPTVVHRDLKSPNLLVDKNWVVKVCDFGLSRMKH
                     HTYLSSKSTAGTPEWMAPEVLRNEPANEKCDVYSSPGVILWELATSRVPWKGLNPMQVV
                     GAVGFQNRLEIPDDIDLTVAQIIRECWQT"
     intron          complement(11638..11710)
                     /gene="F14F18_20"
                     /number=1
     exon            complement(11711..11760)
                     /gene="F14F18_20"
                     /number=2
     intron          complement(11761..11865)
                     /gene="F14F18_20"
                     /number=2
     exon            complement(11866..12033)
                     /gene="F14F18_20"
                     /number=3
     intron          complement(12034..12132)
                     /gene="F14F18_20"
                     /number=3
     exon            complement(12133..12211)
                     /gene="F14F18_20"
                     /number=4
     intron          complement(12212..12406)
                     /gene="F14F18_20"
                     /number=4
     exon            complement(12407..12507)
                     /gene="F14F18_20"
                     /number=5
     intron          complement(12508..12605)
                     /gene="F14F18_20"
                     /number=5
     exon            complement(12606..12674)
                     /gene="F14F18_20"
                     /number=6
     intron          complement(12675..12822)
                     /gene="F14F18_20"
                     /number=6
     exon            complement(12823..12863)
                     /gene="F14F18_20"
                     /number=7
     intron          complement(12864..13088)
                     /gene="F14F18_20"
                     /number=7
     exon            complement(13089..13356)
                     /gene="F14F18_20"
                     /number=8
     intron          complement(13357..13437)
                     /gene="F14F18_20"
                     /number=8
     exon            complement(13438..14056)
                     /gene="F14F18_20"
                     /number=9
     intron          complement(14057..14133)
                     /gene="F14F18_20"
                     /number=9
     exon            complement(14134..14237)
                     /gene="F14F18_20"
                     /number=10
     intron          complement(14238..14415)
                     /gene="F14F18_20"
                     /number=10
     exon            complement(14416..14856)
                     /gene="F14F18_20"
                     /number=11
     intron          complement(14857..14952)
                     /gene="F14F18_20"
                     /number=11
     exon            complement(14953..15048)
                     /gene="F14F18_20"
```

```
RESULT 31
LOCUS       AL159985   114521 bp    DNA             PRI       14-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-22L21 on chromosome 6, complete
            sequence.
ACCESSION   AL159985
VERSION     AL159985.11  GI:11875873
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 114521)
  AUTHORS   Bray-Allen,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  REMARK    Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 17, 2000 this sequence version replaced gi:11340244.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            requests: clonerequest@sanger.ac.uk
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
```

```
Db   90372  ATTTGGAAGAGTTTCAGACC  90391
            ||||||||||||||||||||
Qy    1898  atttggaagagtttcagacc  1917

Query Match              0.8%;  Score 20; DB 13;  Length 103638;
Best Local Similarity  100.0%;  Pred. No. 24;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
intron           complement(15049..15121)
                 /number=12
exon             complement(15122..15565)
                 /gene="F14F18_20"
                 /number=13
intron           16565..17936
                 /gene="F14F18_30"
                 /number=13
gene             complement(join(16565..16786,17110..17447,17579..17936))
                 /gene="F14F18_30"
CDS              complement(join(16565..16786,17110..17447,17579..17936))
                 /gene="F14F18_30"
                 /note="similarity to NLI-interacting factor 1 isoform T1,
                 Gallus gallus, EMBL:AF189773"
                 /codon_start=1
                 /product="putative protein"
                 /protein_id="CAB87659.1"
                 /db_xref="GI:7573353"
                 /translation="MEDDDSPSSRDLDAQNPYDRLLALDTSTVDPNCNLDSVSAIYLA
                 MKSSKLECVDERGQDSLITSVCMEDEDEDEDLDEFDPYLFIKNLPNLSSVPFPRPVLL
                 PKQTRSCPPISLVLDLDETLVHSTLEPCGEVDFTPVNFNEEEHMYYVRCRPHLKEEM
                 ERVSRLFEIIFTASQSIYAEQLLNVLDPKRKLFRHRVYRDSCVFDGNYIKDLSVLG
                 RDLSRVTIVDNSPQAFGFQVENGVPIESWFNDPSDKELLHLLPFLESLIGVEDVRPMI
                 AKFKNLREKIDAAVAAPEYPAEAGDPFER"
exon             complement(16565..16786)
                 /gene="F14F18_30"
                 /number=1
intron           complement(16787..17109)
                 /number=1
exon             complement(17110..17447)
                 /gene="F14F18_30"
                 /number=2
intron           complement(17448..17578)
                 /gene="F14F18_30"
```

```
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP. Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-22L21 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-22L21 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP1-21F7 is at 114422 in this sequence.
            The true right end of clone RP3-417022 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1..114521
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-22L21"
                     /clone_lib="RPCI-11.1"
     repeat_region   123..194
                     /note="118 copies 4 mer tatt 69% conserved"
     repeat_region   865..1152
                     /note="AluJb repeat: matches 192..298 of consensus"
     repeat_region   1953..2008
                     /note="AluSp repeat: matches 1..300 of consensus"
     repeat_region   2185..2417
                     /note="L2 repeat: matches 2700..2750 of consensus"
     repeat_region   3893..3999
                     /note="AluY repeat: matches 84..306 of consensus"
     repeat_region   4165..4428
                     /note="HERVL repeat: matches 3403..3915 of consensus"
     repeat_region   4785..5147
                     /note="HAL1 repeat: matches 1435..1715 of consensus"
     repeat_region   5241..6271
                     /note="MSTB repeat: matches 1..388 of consensus"
     repeat_region   6270..6773
                     /note="HERVL repeat: matches 2126..3260 of consensus"
     repeat_region   6772..7154
                     /note="HERVL repeat: matches 3403..3915 of consensus"
     repeat_region   7463..7668
                     /note="HERVL repeat: matches 4024..4417 of consensus"
     repeat_region   8438..8627
                     /note="MIR repeat: matches 1..224 of consensus"
     repeat_region   9126..9173
                     /note="L1PA3 repeat: matches 5957..6146 of consensus"
     repeat_region   9399..9560
                     /note="24 copies 2 mer ta 77% conserved"
     repeat_region   9872..10167
                     /note="MIR repeat: matches 34..209 of consensus"
     repeat_region   10331..10382
                     /note="AluSx repeat: matches 5..301 of consensus"
     repeat_region   11490..11551
                     /note="13 copies 4 mer tatg 75% conserved"
     repeat_region   11492..11547
                     /note="31 copies 2 mer tatg 75% conserved"
     repeat_region   12339..12357
                     /note="14 copies 4 mer aaat 98% conserved"
     repeat_region   12358..13073
                     /note="L1 repeat: matches 2900..2915 of consensus"
     repeat_region   13074..13395
                     /note="L1MA4 repeat: matches 5576..6300 of consensus"
     repeat_region   13853..13946
                     /note="L1 repeat: matches 2536..2900 of consensus"
     repeat_region   14079..14267
                     /note="MIR repeat: matches 45..138 of consensus"
     repeat_region   14995..15404
                     /note="L1MA10 repeat: matches 5912..6315 of consensus"
     repeat_region   ...
                     /note="MIR repeat: matches 2..207 of consensus"
```

```
     repeat_region   16066..16225
                     /note="L1ME repeat: matches 420..587 of consensus"
     repeat_region   16706..17508
                     /note="L1ME3 repeat: matches 5198..6013 of consensus"
     repeat_region   17677..17818
                     /note="L1 repeat: matches 4558..4701 of consensus"
     repeat_region   17834..18129
                     /note="AluSq repeat: matches 1..295 of consensus"
     repeat_region   18378..19341
                     /note="L1PA8 repeat: matches 5187..6163 of consensus"
     repeat_region   19343..19406
                     /note="32 copies 2 mer aa 76% conserved"
     repeat_region   19407..19665
                     /note="AluY repeat: matches 53..311 of consensus"
     repeat_region   20762..20885
                     /note="MIR repeat: matches 113..232 of consensus"
     repeat_region   22524..23654
                     /note="L1M4 repeat: matches 2927..2945 of consensus"
     repeat_region   25524..25639
                     /note="trigger3b repeat: matches 1..1203 of consensus"
     repeat_region   25639..25680
                     /note="MER5A repeat: matches 1..530 of consensus"
     repeat_region   26701..26842
                     /note="21 copies 2 mer aa 76% conserved"
     repeat_region   27744..28097
                     /note="2 copies 2 mer aa 60% conserved"
     repeat_region   28204..28331
                     /note="71 copies 2 mer aa 60% conserved"
     repeat_region   28716..29064
                     /note="MLT1B repeat: matches 1..554 of consensus"
     repeat_region   29065..29367
                     /note="64 copies 2 mer aa 60% conserved"
     repeat_region   29368..29387
                     /note="L1M4 repeat: matches 2945..2945 of consensus"
     repeat_region   29545..29596
                     /note="L1M4 repeat: matches 2927..2945 of consensus"
     repeat_region   31279..31589
                     /note="L1M4 repeat: matches 5810..5950 of consensus"
     repeat_region   31911..32226
                     /note="AluSq repeat: matches 1..313 of consensus"
     repeat_region   32289..32452
                     /note="AluJb repeat: matches 1..307 of consensus"
     repeat_region   36609..37560
                     /note="L1PA13 repeat: matches 5992..6156 of consensus"
     misc_feature    38216..38375
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   38751..38782
                     /note="MER91a repeat: matches 3..173 of consensus"
     repeat_region   40777..40946
                     /note="16 copies 2 mer ca 100% conserved"
     repeat_region   41788..42159
                     /note="MER5B repeat: matches 1..172 of consensus"
     repeat_region   43207..43513
                     /note="L1MC5 repeat: matches 7311..7699 of consensus"
     repeat_region   43461..43764
                     /note="AluY repeat: matches 1..303 of consensus"
     repeat_region   44458..44847
                     /note="FLAM_C repeat: matches 1..125 of consensus"
     repeat_region   44848..48590
                     /note="L1PA15 repeat: matches 1396..5414 of consensus"
     repeat_region   48448..48590
                     /note="AluJb repeat: matches 1..150 of consensus"
     repeat_region   48851..49074
                     /note="AluUb repeat: matches 1..306 of consensus"
     repeat_region   48891..49074
                     /note="AluSx repeat: matches 1..310 of consensus"
     repeat_region   49075..49819
                     /note="L1PA15 repeat: matches 5414..6151 of consensus"
     repeat_region   50221..50406
                     /note="L2 repeat: matches 2529..2696 of consensus"
     repeat_region   50407..50716
                     /note="AluY repeat: matches 1..304 of consensus"
     repeat_region   50717..50823
                     /note="L2 repeat: matches 2433..2529 of consensus"
     repeat_region   50824..51118
                     /note="AluYa5 repeat: matches 1..299 of consensus"
     repeat_region   51119..51171
                     /note="L2 repeat: matches 2270..2433 of consensus"
     repeat_region   52154..52578
                     /note="MSTB repeat: matches 1..426 of consensus"
     repeat_region   52844..53363
                     /note="MER44B repeat: matches 1..530 of consensus"
     repeat_region   54572..54751
                     /note="MER5A repeat: matches 4..186 of consensus"
     repeat_region   55699..56343
                     /note="L1PB2 repeat: matches 5493..6152 of consensus"
     repeat_region   58368..58536
                     /note="THE1B repeat: matches 1..165 of consensus"
     repeat_region   58537..58827
                     /note="AluSc repeat: matches 1..291 of consensus"
     repeat_region   58828..59013
                     /note="THE1B repeat: matches 165..364 of consensus"
     repeat_region   59014..60590
                     /note="THE1B-INTERNAL repeat: matches 1..1580 of
                     consensus"
     repeat_region   60591..60668
                     /note="THE1B repeat: matches 1..93 of consensus"
     repeat_region   60670..60792
                     /note="THE1-INTERNAL repeat: matches 1060..1182 of
                     consensus"
```

```
Qy      2272 ccaacaatgcccaagctgat 2291
             ||||||||||||||||||||
Db     47417 CCAACAATGCCCAAGCTGAT 47436

Query Match         100.0%;  Score 20;  DB 90;  Length 114521;
Best Local Similarity   100.0%;  Pred. No. 25;
Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;


RESULT 32
HS173D1
HS173D1/c
LOCUS       HS173D1      117338 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from clone 173D1 on chromosome 1p36.21-36.33.
            Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL031984
VERSION     AL031984.13  GI:4678432
KEYWORDS    HTG; CpG Island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117338)
  AUTHORS   Moore,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 24, 1999 this sequence version replaced gi:4547747.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL
            This sequence is the entire insert of clone 173D1. This sequence
            has been finished according to sequence map criteria as follows: An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
```

```
          Mapping Group. Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr1
          173D1 is from the library RPCI1 constructed at the Roswell Park
          Cancer Institute by the group of Pieter de Jong. For further
          details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.

FEATURES             Location/Qualifiers
source               1..117338
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.21-36.33"
                     /clone="RP1-173D1"
                     /clone_lib="RPCI-1"
repeat_region        91..142
                     /note="13 copies 4 mer aggg 75% conserved"
repeat_region        585..992
                     /note="17 copies 24 mer 97% conserved"
repeat_region        1728..1903
                     /note="MER45 repeat: matches 1..178 of consensus"
repeat_region        1909..2078
                     /note="MIR repeat: matches 28..187 of consensus"
repeat_region        4058..4357
                     /note="AluY repeat: matches 1..298 of consensus"
repeat_region        4464..4571
                     /note="MIR repeat: matches 21..151 of consensus"
repeat_region        4726..4844
                     /note="2 copies 54 mer 96% conserved"
repeat_region        5494..5786
                     /note="AluSg repeat: matches 1..292 of consensus"
repeat_region        6503..6807
                     /note="AluJo repeat: matches 1..311 of consensus"
repeat_region        6809..6857
                     /note="MIR repeat: matches 101..150 of consensus"
repeat_region        6858..7181
                     /note="AluSx repeat: matches 1..311 of consensus"
repeat_region        7182..7299
                     /note="MIR repeat: matches 150..262 of consensus"
repeat_region        8932..8993
                     /note="MER53 repeat: matches 126..189 of consensus"
repeat_region        8994..9305
                     /note="AluSq repeat: matches 1..311 of consensus"
repeat_region        9306..9440
                     /note="MER53 repeat: matches 1..126 of consensus"
repeat_region        10094..10654
                     /note="L2 repeat: matches 2152..2746 of consensus"
repeat_region        13515..13745
                     /note="L2 repeat: matches 1886..2084 of consensus"
repeat_region        13746..14051
                     /note="AluSx repeat: matches 1..306 of consensus"
repeat_region        14052..14530
                     /note="L2 repeat: matches 2084..2680 of consensus"
repeat_region        14561..14871
                     /note="AluSx repeat: matches 1..311 of consensus"
repeat_region        16630..16703
                     /note="MIR repeat: matches 101..176 of consensus"
misc_feature         complement(17036..17418)
                     /note="match: STS G11038"
repeat_region        19534..19676
                     /note="MIR repeat: matches 26..168 of consensus"
repeat_region        19811..20021
                     /note="L2 repeat: matches 2521..2705 of consensus"
repeat_region        20103..20402
                     /note="AluSx repeat: matches 1..300 of consensus"
repeat_region        20551..20620
                     /note="L2 repeat: matches 2632..2701 of consensus"
repeat_region        20681..20878
                     /note="MIR repeat: matches 28..238 of consensus"
repeat_region        22445..22468
                     /note="12 copies 2 mer ca 100% conserved"
repeat_region        22496..22714
                     /note="MIR repeat: matches 2..223 of consensus"
repeat_region        22969..23143
                     /note="MIR repeat: matches 71..232 of consensus"
repeat_region        23333..23460
                     /note="2 copies 64 mer 98% conserved"
repeat_region        24447..24757
                     /note="AluSg repeat: matches 1..309 of consensus"
repeat_region        25099..25400
                     /note="AluSx repeat: matches 1..303 of consensus"
repeat_region        25523..25637
                     /note="Charlie5 repeat: matches 5..121 of consensus"
repeat_region        25767..26056
                     /note="AluSx repeat: matches 1..291 of consensus"
repeat_region        26059..26270
                     /note="AluJb repeat: matches 1..312 of consensus"
repeat_region        26313..26624
                     /note="MIR repeat: matches 35..256 of consensus"
repeat_region        27898..27939
                     /note="7 copies 6 mer tcctcc 95% conserved"
repeat_region        29266..29305
                     /note="10 copies 4 mer aagg 80% conserved"
misc_feature         31618..31978
                     /note="match: EST AI350672"
repeat_region        33293..33452
                     /note="MIR repeat: matches 48..203 of consensus"
repeat_region        33832..34011
                     /note="3 copies 64 mer 72% conserved"
repeat_region        33832..33999
                     /note="3 copies 54 mer 74% conserved"
repeat_region        34018..34503
                     /note="9 copies 54 mer 67% conserved"
repeat_region        34018..34497
                     /note="20 copies 24 mer 66% conserved"
repeat_region        34018..34505
                     /note="122 copies 4 mer caca 65% conserved"
repeat_region        34018..34491
                     /note="237 copies 2 mer caca 66% conserved"
repeat_region        34025..34504
                     /note="48 copies 10 mer acacacacac 65% conserved"
repeat_region        34028..34507
                     /note="80 copies 6 mer cacaca 65% conserved"
repeat_region        34065..34512
                     /note="7 copies 64 mer 70% conserved"
repeat_region        34501..34914
                     /note="3 copies 138 mer 70% conserved"
repeat_region        34508..34837
                     /note="33 copies 10 mer acacacacac 66% conserved"
repeat_region        34510..34839
                     /note="55 copies 6 mer acacac 66% conserved"
repeat_region        34511..34846
                     /note="14 copies 24 mer 66% conserved"
repeat_region        34519..34842
                     /note="6 copies 54 mer 67% conserved"
repeat_region        34519..34838
                     /note="160 copies 2 mer ca 66% conserved"
repeat_region        35512..35541
                     /note="15 copies 2 mer tt 90% conserved"
misc_feature         complement(36540..36839)
                     /note="match: EST AA295567"
repeat_region        36962..37180
                     /note="MIR repeat: matches 9..251 of consensus"
repeat_region        37142..37185
                     /note="L2 repeat: matches 2655..2698 of consensus"
repeat_region        37989..38163
                     /note="MIR repeat: matches 47..236 of consensus"
repeat_region        39487..39536
                     /note="5 copies 10 mer gtgtgtgtgt 90% conserved"
repeat_region        39487..39528
                     /note="21 copies 2 mer gt 98% conserved"
repeat_region        39487..39534
                     /note="12 copies 4 mer gtgt 92% conserved"
repeat_region        39831..40133
```

```
repeat_region   40195..40373
                /note="AluSx repeat: matches 1..304 of consensus"
repeat_region   40574..40710
                /note="MIR repeat: matches 32..222 of consensus"
repeat_region   40684..40748
                /note="MER3 repeat: matches 9..149 of consensus"
misc_feature    40795..41105
                /note="Charlie5 repeat: matches 1..62 of consensus"
repeat_region   complement(41047..41397)
                /note="AluSx repeat: matches 1..309 of consensus"
repeat_region   41244..41387
                /note="match: STS L31008"
repeat_region   41263..41538
                /note="72 copies 2 mer tc 63% conserved"
repeat_region   41264..41527
                /note="46 copies 6 mer ttcttt 57% conserved"
repeat_region   41271..41465
                /note="11 copies 24 mer 61% conserved"
repeat_region   41271..41465
                /note="5 copies 39 mer 73% conserved"

RESULT 33
AC022835
Db  68061  CAAGATTCAAACCCAGCTCA  68042
           ||||||||||||||||||||
Qy  1149   caagattcaaacccagctca  1168

Query Match    0.8%;  Score 20;  DB 92;  Length 117338;
Best Local Similarity  100.0%;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Pred. No. 25;

ACCESSION   AC022835
VERSION     AC022835.2  GI:7767812
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-155L3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,J., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassilliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 10, 2000 this sequence version replaced gi:6922258.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----------------------------------------------------------
            Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
```

```
*    24378    25142: contig of 765 bp in length
*    25143    25242: gap of 100 bp
*    25243    26013: contig of 771 bp in length
*    26014    26113: gap of 100 bp
*    26114    26886: contig of 773 bp in length
*    26887    26986: gap of 100 bp
*    26987    27736: contig of 750 bp in length
*    27737    27836: gap of 100 bp
*    27837    28594: contig of 758 bp in length
*    28595    28694: gap of 100 bp
*    28695    29481: contig of 787 bp in length
*    29482    29581: gap of 100 bp
*    29582    30356: contig of 775 bp in length
*    30357    30456: gap of 100 bp
*    30457    31224: contig of 768 bp in length
*    31225    31324: gap of 100 bp
*    31325    32094: contig of 770 bp in length
*    32095    32194: gap of 100 bp
*    32195    32984: contig of 790 bp in length
*    32985    33084: gap of 100 bp
*    33085    33851: contig of 767 bp in length
*    33852    33951: gap of 100 bp
*    33952    34708: contig of 757 bp in length
*    34709    34808: gap of 100 bp
*    34809    35590: contig of 782 bp in length
*    35591    35690: gap of 100 bp
*    35691    36455: contig of 765 bp in length
*    36456    36555: gap of 100 bp
*    36556    37323: contig of 768 bp in length
*    37324    37423: gap of 100 bp
*    37424    38185: contig of 762 bp in length
*    38186    38285: gap of 100 bp
*    38286    39061: contig of 776 bp in length
*    39062    39161: gap of 100 bp
*    39162    39930: contig of 769 bp in length
*    39931    40030: gap of 100 bp
*    40031    40787: contig of 757 bp in length
*    40788    40887: gap of 100 bp
*    40888    41662: contig of 775 bp in length
*    41663    41762: gap of 100 bp
*    41763    42512: contig of 750 bp in length
*    42513    42612: gap of 100 bp
*    42613    43386: contig of 774 bp in length
*    43387    43486: gap of 100 bp
*    43487    44254: contig of 768 bp in length
*    44255    44354: gap of 100 bp
*    44355    45116: contig of 762 bp in length
*    45117    45216: gap of 100 bp
*    45217    45977: contig of 761 bp in length
*    45978    46077: gap of 100 bp
*    46078    46847: contig of 770 bp in length
*    46848    46947: gap of 100 bp
*    46948    47712: contig of 765 bp in length
*    47713    47812: gap of 100 bp
*    47813    48582: contig of 770 bp in length
*    48583    48682: gap of 100 bp
*    48683    49462: contig of 780 bp in length
*    49463    49562: gap of 100 bp
*    49563    50335: contig of 773 bp in length
*    50336    50435: gap of 100 bp
*    50436    51202: contig of 767 bp in length
*    51203    51302: gap of 100 bp
*    51303    52119: contig of 817 bp in length
*    52120    52219: gap of 100 bp
*    52220    53062: contig of 843 bp in length
*    53063    53162: gap of 100 bp
*    53163    53995: contig of 833 bp in length
*    53996    54095: gap of 100 bp
*    54096    54925: contig of 830 bp in length
*    54926    55025: gap of 100 bp
*    55026    55884: contig of 859 bp in length
*    55885    55984: gap of 100 bp
*    55985    56818: contig of 834 bp in length
*    56819    56918: gap of 100 bp
*    56919    57764: contig of 846 bp in length
*    57765    57864: gap of 100 bp
*    57865    58710: contig of 846 bp in length
*    58711    58810: gap of 100 bp
*    58811    59647: contig of 837 bp in length
*    59648    59747: gap of 100 bp
*    59748    60568: contig of 821 bp in length
*    60569    60668: gap of 100 bp
*    60669    61501: contig of 833 bp in length
*    61502    61601: gap of 100 bp
*    61602    62435: contig of 834 bp in length
```

```
Qy      2447 aggagccacaggccaagaag 2466
             ||||||||||||||||||||
Db     69843 AGGAGCCACAGGCCAAGAAG 69862

Query Match             100.0%;  Score 20;  DB 67;  Length 127953;
Best Local Similarity   100.0%;  Pred. No. 25;
Matches   20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT 34
AC027795/c
LOCUS       AC027795   148508 bp   DNA   HTG   03-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-235B15 map 18, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC027795
VERSION     AC027795.2  GI:7684508
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148508)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-235B15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148508)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,J., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Larocque,K., Lamazares,R., Landers,T., Lenoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., Mcpheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 3, 2000 this sequence version replaced gi:7382637.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
```

```
Center project name: L9034
Center clone name: 235_B_15
--------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 138322 bases at least Q40
Consensus quality: 143906 bases at least Q30
Consensus quality: 145997 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 146908; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
---------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*       1       1262: contig of 1262 bp in length
*    1263       1362: gap of     100 bp in length
*    1363       2466: contig of 1104 bp in length
*    2467       2566: gap of     100 bp in length
*    2567       3867: contig of 1301 bp in length
*    3868       3967: gap of     100 bp in length
*    3968       4372: contig of 405 bp in length
*    4373       4472: gap of     100 bp in length
*    4473       7936: contig of 3464 bp in length
*    7937       8036: gap of     100 bp in length
*    8037      11973: contig of 3937 bp in length
*   11974      12073: gap of     100 bp in length
*   12074      16804: contig of 4731 bp in length
*   16805      16904: gap of     100 bp in length
*   16905      22543: contig of 5639 bp in length
*   22544      22643: gap of     100 bp in length
*   22644      27921: contig of 5278 bp in length
*   27922      28021: gap of     100 bp in length
*   28022      33883: contig of 5862 bp in length
*   33884      33983: gap of     100 bp in length
*   33984      39792: contig of 5809 bp in length
*   39793      39892: gap of     100 bp in length
*   39893      49448: contig of 9556 bp in length
*   49449      49548: gap of     100 bp in length
*   49549      59676: contig of 10029 bp in length
*   59677      59776: gap of     100 bp in length
*   59777      69843: contig of 10167 bp in length
*   69844      69943: gap of     100 bp in length
*   69944      95608: contig of 25665 bp in length
*   95609      95708: gap of     100 bp in length
*   95709     120154: contig of 24446 bp in length
*  120155     120254: gap of     100 bp in length
*  120255     148507: contig of 28253 bp in length
*  148508     148508: gap of     100 bp in length.
FEATURES             Location/Qualifiers
   source          1..148508
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="18"
                   /map="18"
                   /clone="RP11-235B15"
                   /clone_lib="RPCI-11 Human Male BAC"
   misc_feature    1..1262
                   /note="assembly_fragment"
   misc_feature    1363..2466
                   /note="assembly_fragment"
   misc_feature    2567..3867
                   /note="assembly_fragment"
   misc_feature    3968..4372
                   /note="assembly_fragment
   misc_feature    clone_end:SP6
                   vector_side:right"
   misc_feature    4473..7936
                   /note="assembly_fragment"
   misc_feature    8037..11973
                   /note="assembly_fragment"
   misc_feature    12074..16804
                   /note="assembly_fragment
   misc_feature    clone_end:T7
                   vector_side:right"
   misc_feature    16905..22543
                   /note="assembly_fragment"
   misc_feature    22644..27921
                   /note="assembly_fragment"
   misc_feature    28022..33883
                   /note="assembly_fragment"
   misc_feature    33984..39792
                   /note="assembly_fragment"
   misc_feature    39893..49448
                   /note="assembly_fragment"
   misc_feature    49549..59577
                   /note="assembly_fragment"
   misc_feature    59678..69844
                   /note="assembly_fragment"
   misc_feature    69945..95609
                   /note="assembly_fragment"
   misc_feature    95710..120155
                   /note="assembly_fragment"
   misc_feature    120256..148508
                   /note="assembly_fragment"
BASE COUNT    44364 a   29387 c   30187 g   42968 t   1602 others
ORIGIN

Qy   1443 agaaatcattctcttggaa 1462
          |||||||||||||||||||||
Db   9577 AGAAATCATCTTCCTTGGAA 9558

Query Match             0.8%;  Score 20;  DB 71;  Length 148508;
Best Local Similarity   100.0%;  Pred. No. 25;
Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

RESULT 35
AC025812
LOCUS          AC025812     149006 bp     DNA                    PRI     03-OCT-2000
DEFINITION     Homo sapiens chromosome 3 clone RP11-182G24, complete sequence.
ACCESSION      AC025812
VERSION        AC025812.11  GI:10518383
KEYWORDS       HTG.
SOURCE         human.
  ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 149006)
  AUTHORS      Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
               Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
               Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
               Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE        Direct Submission
  JOURNAL      Unpublished
REFERENCE      2  (bases 1 to 149006)
  AUTHORS      Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
               Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
               Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
               Yu,S. and Davis,R.W.
  TITLE        Direct Submission
  JOURNAL      Submitted (15-MAR-2000) DNA Sequencing and Technology Center,
               Stanford University, 855 California Avenue, Palo Alto, CA 94304,
               USA
REFERENCE      3  (bases 1 to 149006)
  AUTHORS      Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
               Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
               Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
               Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE        Direct Submission
```

JOURNAL   Submitted (03-OCT-2000) DNA Sequencing and Technology Center,
           Stanford University, 855 California Avenue, Palo Alto, CA 94304,
           USA
COMMENT   On Oct 3, 2000 this sequence version replaced gi:10440608.

           Center: Stanford DNA Sequencing and Technology Development
           Center
           Center code: SDSTDC
           Web site: http://sequence-www.stanford.edu/group/human/
           Contact: hum-info@sequence.stanford.edu

           ---- Project Information
           Center project name: 861
           Center clone name: RP11-182G24

           ---- Summary
           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded or sequenced with an alternate
           chemistry or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

FEATURES   Location/Qualifiers
     source   1..149006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-182G24"
                     /clone_lib="RPCI human BAC library 11"

BASE COUNT   48993 a 29011 c 27339 g 43663 t
ORIGIN

Qy   1110 gaatgagaactgtgcctcag 1129
          |||||||||||||||||||||
Db 120339 GAATGAGAACTGTGCCTCAG 120358

Query Match   0.8%;   Score 20; DB 88; Length 149006;
Best Local Similarity 100.0%;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 36
LOCUS   AC015525
ACCESSION   AC015525
LOCUS   AC015525   149180 bp   DNA   HTG   12-MAR-2000
DEFINITION Homo sapiens chromosome 15 clone RP11-110C15 map 15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC015525
VERSION   AC015525.3 GI:7108019
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 149180)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens chromosome 15, clone RP11-110C15
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 149180)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslaviy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dearellano,K., Devar,K., Donino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,D., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Tirrell,A., Zimmer,A. and Zody,M.

TITLE
JOURNAL   Direct Submission
           Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 28, 2000 this sequence version replaced gi:6563530.
           All repeats were identified using RepeatMasker;
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ------
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu

           ---- Project Information
           Center project name: L4276
           Center clone name: 110_C_15

           ---- Summary Statistics
           Sequencing vector: M13; M77815; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 132714 bases at least Q40
           Consensus quality: 142719 bases at least Q30
           Consensus quality: 146304 bases at least Q20
           Insert size: 156000; agarose-fp
           Insert size: 148380; sum-of-contigs
           Quality coverage: 5.2 in Q20 bases; agarose-fp
           Quality coverage: 5.5 in Q20 bases; sum-of-contigs
           ------
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 9 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           * 1   2388: contig of 2388 bp in length
           * 2389   2488: gap of 100 bp
           * 2489   8437: contig of 5949 bp in length
           * 8438   8537: gap of 100 bp
           * 8538   14871: contig of 6334 bp in length
           * 14872   14971: gap of 100 bp
           * 14972   22311: contig of 7340 bp in length
           * 22312   22411: gap of 100 bp
           * 22412   30639: contig of 8228 bp in length
           * 30640   30739: gap of 100 bp
           * 30740   40297: contig of 9558 bp in length
           * 40298   40397: gap of 100 bp
           * 40398   59678: contig of 19281 bp in length
           * 59679   59778: gap of 100 bp
           * 59779 102333: contig of 42555 bp in length
           * 102334 102433: gap of 100 bp
           * 102434 149180: contig of 46747 bp in length

FEATURES   Location/Qualifiers
     source   1..149180
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-110C15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..2388
                     /note="assembly_fragment"
     misc_feature   2489..8437
                     /note="assembly_fragment"
     misc_feature   8538..14871
                     /note="assembly_fragment"
     misc_feature   14972..22311
                     /note="assembly_fragment"
     misc_feature   22412..30639
                     /note="assembly_fragment"
     misc_feature   30740..40297
                     /note="assembly_fragment"
                     /note="assembly_fragment"
                     clone_end:T7

```
                misc_feature     vector_side:right"
                                 40398..59678
                                 /note="assembly_fragment"
                misc_feature     59779..102333
                                 /note="assembly_fragment
                                 clone_end:SP6
                                 vector_side:right"
                misc_feature     102434..149180
                                 /note="assembly_fragment"
BASE COUNT        42836 a    32086 c   31253 g   42201 t        804 others
ORIGIN

Qy   1588 gaccaggtggacaggtcct 1607
          ||||||||||||||||||||
Db 135751 GACCAGGTGGACAGGGTCCT 135770

RESULT   37
AC069420/c
LOCUS       150683 bp    DNA              HTG       11-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-683J13, WORKING DRAFT
            SEQUENCE, 23 unordered pieces.
ACCESSION   AC069420
VERSION     AC069420.6  GI:10047548
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150683)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,J., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
```

```
Query Match
Best Local Similarity  100.0%;  Score 20;  DB 63;  Length 149180;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
Pred. No. 25;
```

```
                              and Gibbs,R.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 150683)
AUTHORS      Worley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (30-MAY-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Sep 10, 2000 this sequence version replaced gi:9795456.
             -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
             -------------- Project Information
               Center project name: HBGN
               Center clone name: RP11-683J13
             -------------- Summary Statistics
               Assembly program: Phrap; version 0.990329
               Consensus quality: 125821 bases at least Q40
               Consensus quality: 138453 bases at least Q30
               Consensus quality: 144027 bases at least Q20
               Estimated insert size: 142714; sum-of-contigs estimation
               Quality coverage: 0x in Q20 bases; agarose-fp estimation
               Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
             --------------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 23 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1  24905: contig of 24905 bp in length
             *   24906  25005: gap of unknown length
             *   25006  38744: contig of 13739 bp in length
             *   38745  38846: gap of unknown length
             *   38845  53622: contig of 14778 bp in length
             *   53622  53723: gap of unknown length
             *   53623  65642: contig of 11920 bp in length
             *   65642  65742: gap of unknown length
             *   65643  73486: contig of 7744 bp in length
             *   73487  73586: gap of unknown length
             *   73587  81893: contig of 8307 bp in length
             *   81893  81994: gap of unknown length
             *   81894  88596: contig of 6703 bp in length
             *   88597  88697: gap of unknown length
             *   88697  97725: contig of 9029 bp in length
             *   97726  97826: gap of unknown length
             *   97826 103924: contig of 5999 bp in length
             *  103924 104024: gap of unknown length
             *  104025 109912: contig of 5888 bp in length
             *  109913 110012: gap of unknown length
             *  110013 116598: contig of 6586 bp in length
             *  116599 116699: gap of unknown length
             *  116699 120672: contig of 3974 bp in length
             *  120672 120772: gap of unknown length
             *  120773 125139: contig of 4367 bp in length
             *  125140 125239: gap of unknown length
             *  125240 128331: contig of 3092 bp in length
             *  128332 128432: gap of unknown length
             *  128432 131187: contig of 2756 bp in length
             *  131188 131287: gap of unknown length
             *  131288 134177: contig of 2890 bp in length
             *  134178 134277: gap of unknown length
             *  134278 137482: contig of 3205 bp in length
             *  137483 137583: gap of unknown length
             *  137583 138864: contig of 1282 bp in length
             *  138865 138964: gap of unknown length
```

```
LOCUS       AC025326   165490 bp    DNA             HTG       16-JUN-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-621L13 map 2, WORKING DRAFT
            SEQUENCE, 29 unordered pieces.
ACCESSION   AC025326
VERSION     AC025326.3  GI:8569175
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165490)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 2, clone RP11-621L13
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 165490)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Theodore,J., Tirrell,J., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:8099801.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
* ----------------- Genome Center
* Center: Whitehead Institute/ MIT Center for Genome Research
* Center code: WIBR
* Web site: http://www-seq.wi.mit.edu
* Contact: sequence_submissions@genome.wi.mit.edu
* ----------------- Project Information
* Center project name: L5990
* Center clone name: 621_L_13
* ----------------- Summary Statistics
* Sequencing vector: M13; M77815; 100% of reads
* Chemistry: Dye-terminator Big Dye; 100% of reads
* Assembly program: Phrap; version 0.960731
* Consensus quality: 150085 bases at least Q40
* Consensus quality: 150085 bases at least Q40
* Consensus quality: 158263 bases at least Q30
* Consensus quality: 161160 bases at least Q20
* Insert size: 190000; agarose-fp
* Insert size: 162690; sum-of-contigs
* Quality coverage: 3.5 in Q20 bases; agarose-fp
* Quality coverage: 4.1 in Q20 bases; sum-of-contigs
* -----------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
*        1   1610: contig of 1610 bp in length
*     1611   1710: gap of 100 bp
*     1711   2825: contig of 1115 bp in length
*     2826   2925: gap of 100 bp
*     2926   3981: contig of 1056 bp in length
*     3982   4081: gap of 100 bp
*     4082   5162: contig of 1081 bp in length
*     5163   5262: gap of 100 bp
*     5263   6695: contig of 1433 bp in length
*     6696   6795: gap of 100 bp
*     6796   8908: contig of 2113 bp in length
*     8909   9008: gap of 100 bp
*     9009  10640: contig of 1632 bp in length
*    10641  10740: gap of 100 bp
*    10741  11821: contig of 1081 bp in length
*    11822  11921: gap of 100 bp
*    11922  13344: contig of 1423 bp in length
*    13345  13444: gap of 100 bp
*    13445  14731: contig of 1287 bp in length
*    14732  14831: gap of 100 bp
*    14832  17046: contig of 2215 bp in length
*    17047  17146: gap of 100 bp
*    17147  19155: contig of 2009 bp in length
*    19156  19255: gap of 100 bp
*    19256  20955: contig of 1700 bp in length
*    20956  21055: gap of 100 bp
*    21055  23798: contig of 2743 bp in length
*    23799  23898: gap of 100 bp
*    23899  25711: contig of 1813 bp in length
*    25712  25811: gap of 100 bp
*    25812  28549: contig of 2738 bp in length
*    28550  28649: gap of 100 bp
*    28650  33094: contig of 4445 bp in length
*    33095  33194: gap of 100 bp
*    33195  38141: contig of 4947 bp in length
*    38142  38241: gap of 100 bp
*    38242  44592: contig of 6351 bp in length
*    44593  44692: gap of 100 bp
*    44693  51368: contig of 6676 bp in length
*    51369  51468: gap of 100 bp
*    51469  58005: contig of 6537 bp in length
*    58006  58105: gap of 100 bp
*    58106  66252: contig of 8147 bp in length
*    66253  66352: gap of 100 bp
*    66353  77118: contig of 10766 bp in length
```

```
* 138965 142546: contig of 3582 bp in length
* 142547 142646: gap of 100 bp in length
* 142647 145006: contig of 2360 bp in length
* 145007 145106: gap of unknown length
* 145107 147639: contig of 2533 bp in length
* 147640 147739: gap of unknown length
* 147739 149420: contig of 1681 bp in length
* 149421 149520: gap of unknown length
* 149520 149521: contig of unknown length
* 149521 150683: contig of 1163 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..150683
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-683J13"
BASE COUNT    40413 a  33700 c  32450 g  41878 t  2242 others
ORIGIN
```

```
RESULT 38
AC025326

Qy  1588 gaccaggtggacaggtcct  1607
         |||||||||||||||||||
Db   878 GACCAGGTGGACAGGGTCCT 859

Best Local Similarity  100.0%;
Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
Query Match  0.8%;  Score 20;  DB 74;  Length 150683;
```

```
*       77119  77218: gap of        100 bp
*       77219  85575: contig of 8357 bp in length
*       85576  85675: gap of        100 bp
*       85676  96950: contig of 11275 bp in length
*       96951  97050: gap of        100 bp
*       97051 106803: contig of 9753 bp in length
*      106804 106903: gap of        100 bp
*      106904 118011: contig of 11108 bp in length
*      118012 118111: gap of        100 bp
*      118112 133047: contig of 14936 bp in length
*      133048 133147: gap of        100 bp
*      133148 165490: contig of 32343 bp in length.

FEATURES             Location/Qualifiers
     source          1..165490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-621L13"
     misc_feature    1..1610
                     /note="assembly_fragment"
     misc_feature    1711..2825
                     /note="assembly_fragment"
     misc_feature    2926..3981
                     /note="assembly_fragment"
     misc_feature    4082..5162
                     /note="assembly_fragment"
     misc_feature    5263..6695
                     /note="assembly_fragment"
     misc_feature    6796..8908
                     /note="assembly_fragment"
     misc_feature    9009..10640
                     /note="assembly_fragment"
     misc_feature    10741..11821
                     /note="assembly_fragment"
     misc_feature    11922..13344
                     /note="assembly_fragment"
     misc_feature    13445..14731
                     /note="assembly_fragment"
     misc_feature    14832..17046
                     /note="assembly_fragment"
     misc_feature    17147..19155
                     /note="assembly_fragment"
     misc_feature    19256..20955
                     /note="assembly_fragment"
     misc_feature    21056..23798
                     /note="assembly_fragment"
     misc_feature    23899..25711
                     /note="assembly_fragment"
                     /clone_end:T7
                     vector_side:right"
     misc_feature    25812..28549
                     /note="assembly_fragment"
     misc_feature    28650..33094
                     /note="assembly_fragment"
     misc_feature    33195..38141
                     /note="assembly_fragment"
     misc_feature    38242..44592
                     /note="assembly_fragment"
     misc_feature    44693..51368
                     /note="assembly_fragment"
     misc_feature    51469..58005
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:right"
     misc_feature    58106..66252
                     /note="assembly_fragment"
     misc_feature    6633..77118
                     /note="assembly_fragment"
     misc_feature    77219..85575
                     /note="assembly_fragment"
     misc_feature    85676..96950
                     /note="assembly_fragment"
     misc_feature    97051..106803
                     /note="assembly_fragment"
     misc_feature    106904..118011
                     /note="assembly_fragment"
     misc_feature    118112..133047
                     /note="assembly_fragment"
     misc_feature    133148..165490
                     /note="assembly_fragment"
```

```
BASE COUNT    41404 a  40156 c  40246 g  40863 t   2821 others
ORIGIN

Qy       2222 agaaagtgggagttgcctt 2241
              ||||||||||||||||||||
Db  163692 AGAAAGTGGGAGTTGCCTTT 163711

Query Match         0.88;    Score 20;   DB 69;   Length 165490;
Best Local Similarity 100.0%;   Pred. No. 25;
Matches   20;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT  39
AC018876/c
LOCUS       AC018876       165493 bp    DNA               HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-289E15, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC018876
VERSION     AC018876.4  GI:7230959
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165493)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165493)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:6850795.

            --------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----------- Project Information ------------
            Center project name: H_NH0289E15
            ---------------- Summary Statistics ----------------
            Sequencing vector: M13; 81%
            Sequencing vector: plasmid; 19%
            Chemistry: Dye-primer ET; 81% of reads
            Chemistry: Dye-terminator Big Dye; 19% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157046 bases at least Q40
            Consensus quality: 160223 bases at least Q30
            Consensus quality: 161942 bases at least Q20
            Insert size: 174000; agarose-fp
            Insert size: 164793; sum-of-contigs
            Quality coverage: 4.36 in Q20 bases; agarose-fp
            Quality coverage: 4.61 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

RESULT 40

```
LOCUS       AL441923   165914 bp   DNA     HTG      07-APR-2001
DEFINITION  Homo sapiens chromosome X clone RP11-104D21 map p11.21-11.23, ***
            SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION   AL441923
VERSION     AL441923.8  GI:13561280
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165914)
  AUTHORS   Brown,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 8, 2001 this sequence version replaced gi:13396654.
            -------------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------------- Project Information
            Center project name: bA104D21
            -------------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 95% of reads
            Chemistry: Dye-terminator ET-amersham; 4% of reads Consensus
            Quality: 164402 bases at least Q40 Consensus
            Quality: 164723 bases at least Q30 Consensus
            Consensus quality: 164922 bases at least Q20
            Insert size: 165514; sum-of-contigs
            Insert size: 156718; 6.1% error; agarose-fp
            Quality coverage: 6.95x in Q20 bases; sum-of-contigs Quality
            coverage: 9.79x in Q20 bases; agarose-fp
            -------------------
FEATURES             Location/Qualifiers
     source          1..165493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-289E15"
     *        1    1322: contig of 1322 bp in length
     *     1323    1422: gap of unknown length
     *     8681    8681: gap of 7259 bp in length
     *     8682    8781: gap of unknown length
     *     8782   28642: contig of 19861 bp in length
     *    28643   28742: gap of unknown length
     *    28644   46484: contig of 17742 bp in length
     *    46485   46584: gap of unknown length
     *    46585   68690: contig of 22106 bp in length
     *    68691   68790: gap of unknown length
     *    68791   99194: contig of 30404 bp in length
     *    99195   99294: gap of unknown length
     *    99295  131431: contig of 32137 bp in length
     *   131432  131531: gap of unknown length
     *   131532  165493: contig of 33962 bp in length
BASE COUNT    52008 a   30658 c   30763 g   51358 t    706 others
ORIGIN
```

Query Match Best Local Similarity 100.0%; Score 20; DB 65; Length 165493;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1716 gcttcaccttgctgtgg 1735
           |||||||||||||||||||
Db   11874 GCTTCACCCTTTGCTGGTGG 11855
```

RESULT 41

```
LOCUS       AL162577   166344 bp   DNA     HTG      23-JAN-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-238B16, *** SEQUENCING IN
            PROGRESS ***, 8 unordered pieces.
ACCESSION   AL162577
VERSION     AL162577.11  GI:9988271
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166344)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 2000 this sequence version replaced gi:9884780.
            -------------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------------- Project Information
FEATURES             Location/Qualifiers
     source          1..165914
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="p11.21-11.23"
                     /clone="RP11-104D21"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..157250
                     /note="assembly_fragment:00834"
     misc_feature    157351..159384
                     /note="assembly_fragment:01928"
     misc_feature    159485..161555
                     /note="assembly_fragment:02600"
     misc_feature    161656..163744
                     /note="assembly_fragment:02933"
     misc_feature    163845..165914
                     /note="assembly_fragment:03272"
                     NOTE: This is a 'working draft' sequence. It currently
                     consists of 5 contigs. The true order of the pieces
                     is not known and their order in this sequence record is
                     arbitrary. Gaps between the contigs are represented as
                     runs of N, but the exact sizes of the gaps are unknown.
                     This record will be updated with the finished sequence
                     as soon as it is available and the accession number will
                     be preserved.
     *      157250: contig of 157250 bp in length
     *   157251  157350: gap of 100 bp in length
     *   157351  159384: contig of 2034 bp in length
     *   159385  159484: gap of 100 bp in length
     *   159485  161555: contig of 2071 bp in length
     *   161556  161655: gap of 100 bp in length
     *   161656  163744: contig of 2089 bp in length
     *   163745  163844: gap of 100 bp in length
     *   163845  165914: contig of 2070 bp in length
BASE COUNT    49534 a   32899 c   33616 g   49465 t    400 others
ORIGIN
```

Query Match Best Local Similarity 100.0%; Score 20; DB 81; Length 165914;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    2247 catgaaggtctgcttggag 2266
           |||||||||||||||||||
Db  137911 CATGAAGGTCTGCTTTGGAG 137892
```

Center project name: ba238B16
----------   Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162011 bases at least Q40
Consensus quality: 163747 bases at least Q30
Consensus quality: 164823 bases at least Q20
Insert size: 165644; sum-of-contigs
Insert size: 171988; 10.5% error; agarose-fp
Quality coverage: 4.90x in Q20 bases; agarose-fp
coverage: 4.99x in Q20 bases; sum-of-contigs Quality

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES   Location/Qualifiers
source   1..166344
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /clone="RP11-238B16"
                /clone_lib="RPCI-11.1"
misc_feature   1..11500
                /note="assembly_fragment:00621
                clone_end:SP6"
                vector_side:left"
misc_feature   11601..35909
                /note="assembly_fragment:01398
                fragment_chain:1"
misc_feature   36010..39789
                /note="assembly_fragment:02316
                fragment_chain:1"
misc_feature   39890..47363
                /note="assembly_fragment:00526"
misc_feature   47464..98169
                /note="assembly_fragment:00873"
misc_feature   98270..110390
                /note="assembly_fragment:00897"
misc_feature   110491..122457
                /note="assembly_fragment:01455"
misc_feature   12258..166344
                /note="assembly_fragment:01966
                clone_end:T7
                vector_side:right"

1   11500: contig of 11500 bp in length
11501   11600: gap of 100 bp in length
11601   35909: contig of 24309 bp in length
35910   36009: gap of 100 bp in length
36010   39789: contig of 3780 bp in length
39790   39889: gap of 100 bp in length
39890   47363: contig of 7474 bp in length
47364   47463: gap of 100 bp in length
47464   98169: contig of 50706 bp in length
98170   98269: gap of 100 bp in length
98270   110390: contig of 12121 bp in length
110391   110490: gap of 100 bp in length
110491   122457: contig of 11967 bp in length
122458   122557: gap of 100 bp in length
122558   166344: contig of 43787 bp in length.

Qy   2272 ccaacaatgcccaagtgat 2291
          ||||||||||||||||| |||
Db 66987 CCAACAATGCCCAAGCTGAT 66968

RESULT 42
Db 66987 CCAACAATGCCCAAGCTGAT 66968

AC021927
AC021927
LOCUS   AC021927   168997 bp   DNA   HTG   03-MAR-2000
DEFINITION Homo sapiens clone RP11-29P22, WORKING DRAFT SEQUENCE, 26 unordered
            pieces.
ACCESSION   AC021927
VERSION   AC021927.3 GI:7144973
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168997)
  Homo sapiens chromosome, clone RP11-29P22
  Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Unpublished
  JOURNAL
REFERENCE   2 (bases 1 to 168997)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 3, 2000 this sequence version replaced gi:6984426.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----
            Center: Whitehead Institute/MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu

            Project Information
            Center project name: L4856
            Center clone name: 29_P_22

            Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 156612 bases at least Q40
            Consensus quality: 162417 bases at least Q30
            Consensus quality: 164820 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 164497; sum-of-contigs
            Quality coverage: 3.6 in Q20 bases; agarose-fp
            Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will

Query Match
Best Local Similarity 100.0%;   0.8%;   Score 20; DB 79; Pred. No. 25;   Length 166344;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT   50249 a   32106 c   31398 g   51875 t   716 others
ORIGIN

* be preserved.

```
FEATURES             Location/Qualifiers
     source          1..168997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="RPCI-11 29P22"
                     /clone="RPCI-11 Human Male BAC"
*    misc_feature    1..1046
                     /note="assembly_fragment"
*    misc_feature    1147..2513
                     /note="assembly_fragment"
*    misc_feature    2614..3854
                     /note="assembly_fragment"
*    misc_feature    3955..5409
                     /note="assembly_fragment"
*    misc_feature    5510..8344
                     /note="assembly_fragment"
*    misc_feature    8445..11304
                     /note="assembly_fragment"
*    misc_feature    11405..14151
                     /note="assembly_fragment"
*    misc_feature    14252..17174
                     /note="assembly_fragment"
```

```
*       1    1046: contig of 1046 bp in length
*    1047    1146: gap of 100 bp
*    1147    2513: contig of 1367 bp in length
*    2514    2613: gap of 100 bp
*    2614    3854: contig of 1241 bp in length
*    3855    3954: gap of 100 bp
*    3955    5409: contig of 1455 bp in length
*    5510    8344: contig of 2835 bp in length
*    8345    8444: gap of 100 bp
*    8445   11304: contig of 2860 bp in length
*   11305   11404: gap of 100 bp
*   11405   14151: contig of 2747 bp in length
*   14152   14251: gap of 100 bp
*   14252   17174: contig of 2923 bp in length
*   17175   17274: gap of 100 bp
*   17275   19939: contig of 2665 bp in length
*   19940   20039: gap of 100 bp
*   20040   23411: contig of 3372 bp in length
*   23412   23511: gap of 100 bp
*   23512   27130: contig of 3619 bp in length
*   27131   27230: gap of 100 bp
*   27231   29770: contig of 2540 bp in length
*   29771   29870: gap of 100 bp
*   29871   36121: contig of 6251 bp in length
*   36122   36221: gap of 100 bp
*   36222   41978: contig of 5757 bp in length
*   41979   42078: gap of 100 bp
*   42079   49159: contig of 7081 bp in length
*   49160   49259: gap of 100 bp
*   49260   57801: contig of 8542 bp in length
*   57802   57901: gap of 100 bp
*   57902   64360: contig of 6459 bp in length
*   64361   64460: gap of 100 bp
*   64461   72524: contig of 8064 bp in length
*   72525   72624: gap of 100 bp
*   72625   81728: contig of 9104 bp in length
*   81729   81828: gap of 100 bp
*   81829   92259: contig of 10431 bp in length
*   92260   92359: gap of 100 bp
*   92360  100902: contig of 8543 bp in length
*  100903  101002: gap of 100 bp
*  101003  110082: contig of 9080 bp in length
*  110083  110182: gap of 100 bp
*  110183  124956: contig of 14774 bp in length
*  124957  125056: gap of 100 bp
*  125057  139552: contig of 14496 bp in length
*  139553  139652: gap of 100 bp
*  139653  153065: contig of 13413 bp in length
*  153066  153165: gap of 100 bp
*  153166  168997: contig of 15832 bp in length.
```

```
     misc_feature    17275..19939
                     /note="assembly_fragment"
     misc_feature    20040..23411
                     /note="assembly_fragment"
     misc_feature    23512..27130
                     /note="assembly_fragment"
     misc_feature    27231..29770
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    29871..36121
                     /note="assembly_fragment"
     misc_feature    36222..41978
                     /note="assembly_fragment"
     misc_feature    42079..49159
                     /note="assembly_fragment"
     misc_feature    49260..57801
                     /note="assembly_fragment"
     misc_feature    57902..64360
                     /note="assembly_fragment"
     misc_feature    64461..72524
                     /note="assembly_fragment"
     misc_feature    72625..81728
                     /note="assembly_fragment"
     misc_feature    81829..92259
                     /note="assembly_fragment"
     misc_feature    92360..100902
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    101003..110082
                     /note="assembly_fragment"
     misc_feature    110183..124956
                     /note="assembly_fragment"
     misc_feature    125057..139552
                     /note="assembly_fragment"
     misc_feature    139653..153065
                     /note="assembly_fragment"
     misc_feature    153166..168997
                     /note="assembly_fragment"
```

```
RESULT  43

ACCESSION    AC025191
LOCUS        AC025191   177868 bp    DNA     HTG      03-MAY-2000
DEFINITION   Homo sapiens chromosome 4 clone RP11-285A15 map 4, WORKING DRAFT
             SEQUENCE, 43 unordered pieces.
ACCESSION    AC025191
VERSION      AC025191.2  GI:7684444
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 177868)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 4, clone RP11-285A15
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 177868)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
```

```
Qy    1443 agaaatcatcttccttggaa 1462
           ||||||||||||||||||||
Db  110794 AGAAATCATCTTCCTTGGAA 110813
```

```
Query Match             0.8%;  Score 20;  DB 66;  Length 168997;
Best Local Similarity   100.0%;  Pred. No. 25;
Matches    20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

BASE COUNT    48830 a   34497 c   33807 g   49359 t   2504 others
ORIGIN
```

```
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Ilev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,R., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Whitehead Institute/ MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 3, 2000 this sequence version replaced gi:7188869.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L5067
            Center clone name: 285_A_15
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 150228 bases at least Q40
            Consensus quality: 163346 bases at least Q30
            Consensus quality: 169409 bases at least Q20
            Insert size: 189000; agarose-fp
            Insert size: 173668; sum-of-contigs
            Quality coverage: 2.9 in Q20 bases; agarose-fp
            Quality coverage: 3.2 in Q20 bases; sum-of-contigs
            ---------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 43 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *          1   1215: contig of 1215 bp in length
          *       1216   1315: gap of 100 bp
          *       1316   2523: contig of 1208 bp in length
          *       2524   2623: gap of 100 bp
          *       2624   3811: contig of 1188 bp in length
          *       3812   3911: gap of 100 bp
          *       3912   5309: contig of 1398 bp in length
          *       5310   5409: gap of 100 bp
          *       5410   7250: contig of 1841 bp in length
          *       7251   7350: gap of 100 bp
          *       7351   8970: contig of 1620 bp in length
          *       8971   9070: gap of 100 bp
          *       9071  10289: contig of 1219 bp in length
          *      10290  10389: gap of 100 bp
          *      10390  11577: contig of 1188 bp in length
          *      11578  11677: gap of 100 bp
          *      11678  12960: contig of 1283 bp in length
          *      12961  13060: gap of 100 bp
          *      13061  14891: contig of 1831 bp in length
          *      14892  14991: gap of 100 bp
          *      14992  16528: contig of 1537 bp in length
          *      16529  16628: gap of 100 bp
          *      16629  18711: contig of 2083 bp in length
          *      18712  18811: gap of 100 bp
          *      18812  20214: contig of 1403 bp in length
          *      20215  20314: gap of 100 bp
          *      20315  22282: contig of 1968 bp in length
          *      22283  22382: gap of 100 bp
          *      22383  24680: contig of 2298 bp in length
          *      24681  24780: gap of 100 bp
          *      24781  27243: contig of 2463 bp in length
          *      27244  27343: gap of 100 bp
          *      27344  29231: contig of 1888 bp in length
          *      29232  29331: gap of 100 bp
          *      29332  32345: contig of 3014 bp in length
          *      32346  32445: gap of 100 bp
          *      32446  35005: contig of 2560 bp in length
          *      35006  35105: gap of 100 bp
          *      35106  37773: contig of 2668 bp in length
          *      37774  37873: gap of 100 bp
          *      37874  40633: contig of 2760 bp in length
          *      40634  40733: gap of 100 bp
          *      40734  43133: contig of 2400 bp in length
          *      43134  43233: gap of 100 bp
          *      43234  45935: contig of 2702 bp in length
          *      45936  46035: gap of 100 bp
          *      46036  48895: contig of 2860 bp in length
          *      48896  48995: gap of 100 bp
          *      48996  51307: contig of 2312 bp in length
          *      51308  51407: gap of 100 bp
          *      51408  55313: contig of 3906 bp in length
          *      55314  55413: gap of 100 bp
          *      55414  58208: contig of 2795 bp in length
          *      58209  58308: gap of 100 bp
          *      58309  62616: contig of 4308 bp in length
          *      62617  62716: gap of 100 bp
          *      62717  67413: contig of 4697 bp in length
          *      67414  67513: gap of 100 bp
          *      67514  70733: contig of 3220 bp in length
          *      70734  70833: gap of 100 bp
          *      70834  75405: contig of 4572 bp in length
          *      75406  75505: gap of 100 bp
          *      75506  80523: contig of 5018 bp in length
          *      80524  80623: gap of 100 bp
          *      80624  86066: contig of 5443 bp in length
          *      86067  86166: gap of 100 bp
          *      86167  92000: contig of 5834 bp in length
          *      92001  92100: gap of 100 bp
          *      92101  96663: contig of 4563 bp in length
          *      96664  96763: gap of 100 bp
          *      96764 102888: contig of 6125 bp in length
          *     102889 102988: gap of 100 bp
          *     102989 110032: contig of 7044 bp in length
          *     110033 110132: gap of 100 bp
          *     110133 117903: contig of 7771 bp in length
          *     117904 118003: gap of 100 bp
          *     118004 125508: contig of 7505 bp in length
          *     125509 125608: gap of 100 bp
          *     125609 137315: contig of 11707 bp in length
          *     137316 137415: gap of 100 bp
          *     137416 148418: contig of 11003 bp in length
          *     148419 148518: gap of 100 bp
          *     148519 161620: contig of 13102 bp in length
          *     161621 161720: gap of 100 bp
          *     161721 177868: contig of 16148 bp in length.
FEATURES             Location/Qualifiers
     source          1..177868
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-285A15"
```

```
Qy    1731 ggtggttgcccaacagc 1750
            |||||||||||||||||
Db   57383 GGTGGTTGCCCCCAACCAGC 57402
```

```
Query Match         0.8%;   Score 20;   DB 69;   Length 177868;
Best Local Similarity 100.0%;   Pred. No. 25;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
```

```
RESULT    44
AP001899/c
LOCUS          AP001899    183412 bp            DNA            HTG          26-JUL-2000
DEFINITION     Homo sapiens chromosome 18 clone RP11-701C7 map 18q12, WORKING
               DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION      AP001899
VERSION        AP001899.3  GI:9501833
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens DNA, clone:RP11-701C7.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 183412)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
               Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Homo sapiens 183,412 genomic DNA of 18q12
JOURNAL        Published Only in DataBase (2000) In press
REFERENCE      2  (bases 1 to 183412)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
               Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Direct Submission
JOURNAL        Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
               (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
               Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT        On Jul 26, 2000 this sequence version replaced gi:8117550.
               -------------------- Genome Center
               Center: RIKEN Genomic Sciences Center(GSC)
               Center code: RIKEN
               Web site: http://hgp.gsc.riken.go.jp/
               Contact: hattori@gsc.riken.go.jp
               -------------- Project Information
               Center project name: HumDraft18
               Center clone name: RP11-701C7
               -------------- Summary Statistics
               Sequencing vector: PCR products; 100% of reads
               Chemistry: Dye-terminator ET-amersham; 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 180400 bases at least Q40
               Consensus quality: 181718 bases at least Q30
               Consensus quality: 182304 bases at least Q20
               Insert size: 182712; sum-of-contigs
               Quality coverage: 10.87x in Q20 bases; sum-of-contigs
               ------------------------
               NOTE: This is a 'working draft' sequence. It currently consists of
               8 contigs. The true order of the pieces is not known and their
               order in this sequence record is arbitrary. Gaps between the
               contigs are represented as runs of N, but the exact sizes of the gaps
               are unknown. This record will be updated with the finished sequence
               as soon as it is available and the accession number will be
               preserved.
               *
               NOTE: This is a 'working draft' sequence. It currently
               * consists of 8 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *         1    42740: contig of 42740 bp in length
               *     42741    42840: gap of        100 bp in length
               *     42841    79938: contig of 37098 bp in length
               *     79939    80038: gap of        100 bp in length
               *     80039   109211: contig of 29173 bp in length
               *    109212   109311: gap of        100 bp in length
               *    109312   135418: contig of 26107 bp in length
               *    135419   135518: gap of        100 bp in length
               *    135519   161816: contig of 26298 bp in length
               *    161817   161916: gap of        100 bp in length
               *    161917   174638: contig of 12722 bp in length
               *    174639   174738: gap of        100 bp in length
               *    174739   180277: contig of  5539 bp in length
               *    180278   180377: gap of        100 bp in length
               *    180378   183412: contig of  3035 bp in length.
FEATURES             Location/Qualifiers
     source          1..183412
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q12"
                     /clone="RP11-701C7"
     misc_feature    1..42740
                     /note="assembly_fragment"
     misc_feature    42841..79938
                     /note="assembly_fragment"
     misc_feature    80039..109211
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    109312..135418
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    135519..161816
                     /note="assembly_fragment"
     misc_feature    1..1215
                     /note="assembly_fragment"
     misc_feature    1316..2523
                     /note="assembly_fragment"
     misc_feature    2624..3811
                     /note="assembly_fragment"
     misc_feature    3912..5309
                     /note="assembly_fragment"
     misc_feature    5410..7250
                     /note="assembly_fragment"
     misc_feature    7351..8970
                     /note="assembly_fragment"
     misc_feature    9071..10289
                     /note="assembly_fragment"
     misc_feature    10390..11577
                     /note="assembly_fragment"
     misc_feature    11678..12960
                     /note="assembly_fragment"
     misc_feature    13061..14691
                     /note="assembly_fragment"
     misc_feature    14992..16528
                     /note="assembly_fragment"
     misc_feature    16629..18711
                     /note="assembly_fragment"
     misc_feature    18812..20214
                     /note="assembly_fragment"
     misc_feature    20315..22282
                     /note="assembly_fragment"
     misc_feature    22383..24680
                     /note="assembly_fragment"
     misc_feature    24781..27243
                     /note="assembly_fragment"
     misc_feature    27344..29231
                     /note="assembly_fragment"
     misc_feature    29332..32345
                     /note="assembly_fragment"
```

```
misc_feature    /note="assembly_fragment"
                16917..174638
misc_feature    /note="assembly_fragment"
                17439..180227
misc_feature    /note="assembly_fragment"
                180378..183412
                /note="assembly_fragment"

BASE COUNT      49688 a    39575 c   41016 g   52433 t    700 others
ORIGIN

Qy   1350 gctgcagcttccaaacttcc 1369
          |||||||||||||||||||||||
Db 121048 GCTGCAGCTTCCCAACTTCC 121029

RESULT  45
ACC009591
LOCUS           AC009591      183855 bp      DNA             HTG          26-MAY-2000
DEFINITION      Homo sapiens chromosome 4 clone RP11-335K21 map 4, WORKING DRAFT
                SEQUENCE, 6 unordered pieces.
ACCESSION       AC009591
VERSION         AC009591.3  GI:7107764
KEYWORDS        HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 183855)
AUTHORS         Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL         Unpublished
REFERENCE       2  (bases 1 to 183855)
AUTHORS         Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
                Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
                Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
                Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
                Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forest,C.,
                Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
                Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
                Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
                Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
                Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
                Naylor,J., Nioff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
                Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
                Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
                Tesfaye,S., Torruella-Miller,I., Vassliiev,H., Vo,A., Wagner,A.,
                Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
JOURNAL         Direct Submission
                Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
                On Feb 28, 2000 this sequence version replaced gi:6094611.
COMMENT         All repeats were identified using RepeatMasker:
                http://ftp.genome.washington.edu/RM/RepeatMasker.html
                ---------------                      Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                ---------------                      Project Information
                Center project name: L1744
                Center clone name: 335_K_21
                ---------------                      Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-primer-amersham; 2% of reads
                Chemistry: Dye-terminator Big Dye; 98% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158308 bases at least Q40
```

```
FEATURES             Location/Qualifiers
     source          1..183855
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-335K21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1031
                     /note="assembly_fragment"
     misc_feature    1132..2356
                     /note="assembly_fragment"
     misc_feature    2457..3536
                     /note="assembly_fragment"
     misc_feature    3637..8344
                     /note="assembly_fragment"
     misc_feature    8445..45297
                     /note="assembly_fragment"
                     /clone_end:I7
                     /vector_side:right"
     misc_feature    45398..183855
                     /note="assembly_fragment
                     /clone_end:SP6
                     /vector_side:left"

BASE COUNT      62884 a    32653 c   31457 g   56345 t    516 others
ORIGIN

Qy   1731 ggtggttggcccccaaccagc 1750
          ||||||||||||||||||||||
Db 61148 GGTGGTTGGCCCCAACCAGC 61167
```

* Consensus quality: 170972 bases at least Q30
* Consensus quality: 177434 bases at least Q20
* Insert size: 190000; agarose-fp
* Insert size: 183355; sum-of-contigs
* Quality coverage: 5.2 in Q20 bases; agarose-fp
* Quality coverage: 5.4 in Q20 bas.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
       1 1031: contig of 1031 bp in length
    1032 1131: gap of        100 bp
    1132 2356: contig of 1225 bp in length
    2357 2456: gap of        100 bp
    2457 3536: contig of 1080 bp in length
    3537 3636: gap of        100 bp
    3637 8344: contig of 4708 bp in length
    8345 8444: gap of        100 bp
    8445 45297: contig of 36853 bp in length
   45298 45397: gap of       100 bp
   45398 183855: contig of 138458 bp in length.
```

```
Query Match          0.8%;   Score 20;   DB 83;   Length 183412;
Best Local Similarity 100.0%;   Pred. No. 25;
Matches 20; Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Query Match          0.8%;   Score 20;   DB 61;   Length 183855;
Best Local Similarity 100.0%;   Pred. No. 25;
Matches 20; Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

Search completed: October 28, 2001, 21:57:56
Job time: 12135 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: October 28, 2001, 19:35:01 ; Search time 163.36 seconds
(without alignments)
9536.141 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgcttgctcgct.........agaaggtcagagccagtga 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 8

Total number of hits satisfying chosen parameters: 939240

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2481	100.0	2958	21	AAA58453	Human prostate can
2	2274	91.7	2478	21	AAA52810	Human sulphatase G
3	1600	64.5	2546	21	AAC76445	Human ORFX ORF2000
4	247	10.0	26664	21	AAA60207	Human prostate can
5	237	9.6	238	21	AAZ80231	Human colon cancer
6	26	1.0	326	21	AAZ80390	Murine prostate ca
7	23	0.9	23	21	AAA60373	Human prostate can
8	21	0.8	143	21	AAC24564	Human secreted pro
9	21	0.8	38	21	AAA60275	Human prostate can
10	20	0.8	308	21	AAF09297	Fusarium venenatum
11	20	0.8	457	21	AAC57035	Pinus radiata tran
c 12	19	0.8	19	21	AAA60211	Human HPC2 CDNA 5'
c 13	19	0.8	19	21	AAA60229	Human HPC2 CDNA se
c 14	19	0.8	19	21	AAA60230	Human HPC2 CDNA se
c 15	19	0.8	19	21	AAA60233	Human HPC2 CDNA se
c 16	19	0.8	19	21	AAA60235	Human HPC2 CDNA se
c 17	19	0.8	19	21	AAA60364	Human HPC2 CDNA ex
c 18	19	0.8	29	21	AAA60214	Human HPC2 CDNA se
c 19	19	0.8	37	21	AAA60362	Human HPC2 CDNA ex
c 20	19	0.8	38	21	AAA60307	Human HPC2 CDNA ex
c 21	19	0.8	526	20	AAV88751	EST clone HK189.
c 22	20	0.8	7642	20	AAV82020	Moraxella catarrha
c 23	18	0.7	18	21	AAA60219	Human HPC2 CDNA se
c 24	18	0.7	18	21	AAA60220	Human HPC2 CDNA se
c 25	18	0.7	18	21	AAA60221	Human HPC2 CDNA se
c 26	18	0.7	18	21	AAA60222	Human HPC2 CDNA se
c 27	18	0.7	18	21	AAA60223	Human HPC2 CDNA se
c 28	18	0.7	18	21	AAA60224	Human HPC2 CDNA se
c 29	18	0.7	18	21	AAA60225	Human HPC2 CDNA se
c 30	18	0.7	18	21	AAA60226	Human HPC2 CDNA se
c 31	18	0.7	18	21	AAA60227	Human HPC2 CDNA se
c 32	18	0.7	18	21	AAA60228	Human HPC2 CDNA se
c 33	18	0.7	18	21	AAA60232	Human HPC2 CDNA se
c 34	18	0.7	18	21	AAA60236	Human HPC2 CDNA se
c 35	18	0.7	18	21	AAA60237	Human HPC2 CDNA se
c 36	18	0.7	18	21	AAA60238	Human HPC2 CDNA se
c 37	18	0.7	18	21	AAA60240	Human HPC2 CDNA se
c 38	18	0.7	18	21	AAA60243	Human HPC2 CDNA se
c 39	18	0.7	45	20.	AAZ33905	Human PRO274 hybri
c 40	20.		45	20.	AAC78611	Human PRO274 hybri
c 41	20	0.7	177	20	AAX18066	Coding sequence fo
c 42	18	0.7	531	18	AAT67773	H. pylori cytoplas
c 43	18	0.7	531	18	AAT77453	H. pylori cytoplas
c 44	18	0.7	726	21	AAZ80749	Human colon cancer
c 45	18	0.7	791	21	AAC44859	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAA58453

ID   AAA58453 standard; cDNA; 2958 BP.
XX
AC   AAA58453;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   51..2531
FT   /*tag= a
FT   /product= "HPC2"
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.

```
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Claim 3; Page 98-100; 157pp; English.
XX

CC   The present sequence is the coding sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;

Query Match             100.0%;  Score 2481;  DB 21;  Length 2958;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2481;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy        1 atgtgggcgcttgctcgctcgctcgcgtccggcgaggacgc 60
            |||||||||||||||||||||||||||||||||||||||||
Db       51 atgtgggcgcttgctcgctcgctcgcgtccggcgaggacgc 110

Qy       61 accaatatcgcaggcacacccgcccgcccgcccgccgtcggcacctg 120
            |||||||||||||||||||||||||||||||||||||||||||||||
Db      111 accaatatcgcaggcacacccgcccgcccgcccgccgtcggcacctg 170

Qy      121 cgcacgcgagagaaagccggacgtcgggcgttcaggagaagcgttacctg 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      171 cgcacgcgagagaaagccggacgtcgggcgttcaggagaagcgttacctg 230

Qy      181 caggtggtgcggggtagccggactcggcgggtgctctacgtcttctccggttc 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      231 caggtggtgcggggtagccggactcggcgggtgctctacgtcttctccggttc 290

Qy      241 aaccggtatctcttcaactgtggagaagcgttcaggagaagcgttacctg 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      291 aaccggtatctcttcaactgtggagaagcgttcatgagagcagcaagtta 350

Qy      301 aaggttgctcgcctggacaacagatattcctgacacgtgctcaattgtgggggc 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      351 aaggttgctcgcctggacaacatattcctgacacgtgctcaattgtgggggc 410

Qy      361 ttaagtggaatgaattcttacttttaaagaaaccagggccttccaaagtgtgtactttctgga 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      411 ttaagtggaatgaattcttacttttaaagaaaccggcttccaaagtgtgtactttctgga 470

Qy      421 cctcccacaactggaaaaatacctcgaagcaatcaaaaatatttctggtccattgaaaga 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      471 cctccacactgtggaaaatacctcgaagcaatcaaaaatatttctggtccattgaaaga 530

Qy      481 atagaactggctgtgcgccccactcctgccccagaatgagaactgtgacagtt 540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      531 atagaactggctgtgcgccccactcctgccccagaatgagaactgtgacagtt 590

Qy      541 taccagaatcccacatacacagtgaacagagagggagggaaagcacaaccatgaaccatggca 600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      591 taccagaatcccacatacacagtgaacagagagggaaagcacaaccatggcaagagtcca 650

Qy      601 gaaaggcctccaagctcagtccagctccagtccgaatgcaaat 660
            |||||||||||||||||||||||||||||||||||||||||||
Db      651 gaaaggcctccaagctcagtccagatcttcagactcgaatgaaaat 710

Qy      661 gagccacacctccacatggttagcacgggactcttccctgtc 720
            ||||||||||||||||||||||||||||||||||||||||||
Db      711 gagccacacctccacatggttagcacgggactcttccctgtc 770

Qy      721 gtagctttcatctgtaagcttcacttaaagagagaaactcttggtgctcaaagcaaag 780
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      771 gtagcttctcatctgtaagcttcacttaaagagagaaaacttcttggtgctcaaagcaaag 830

Qy      781 gagatgggcttccagtgggacagctgccatcgtcccatcattgctgctgtcaaagac 840
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      831 gagatgggcttccagtgggacagctgccatcgtcccatcattgctgctgtcaaagac 890

Qy      841 gggaaaaagcatcactcatgaatgaacagactgccatcctctccatcatttgctcctcca 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      891 gggaaaaagcatcactcatggaagaatctctgaaatgagatgaagagaagaaggatcctcca 950

Qy      901 gatcctgggctgctgctttgtggttggtagaatgtccagatgaaagcttcattcaaccatc 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      951 gatcctgggctgctgctttgtggttggtagaatgtccagatgaaagcttcattcaaccatc 1010

Qy      961 tgtgagaatgccacctttcagaggttaccaaggaaggcagagtgcccctggccttgtg 1020
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1011 tgtgagaatgccacctttcagaggttaccaaggaaggcagagtgcccctggccttgtg 1070

Qy     1021 gttcacatggccccagacactgtgcttgtggacaagaagcaagtcagtggaattgaggg 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1071 gttcacatggccccagacactgtgcttgtggacaagaagcaagtcagtggaattgaggg 1130

Qy     1081 tttggcctgacaccagacacttggtcctgaatgaacaactgttgctcagttcacaacctt 1140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1131 tttggcctgacaccagacacttggtcctgaatgaacaactgttgctcagttcacaacctt 1190

Qy     1141 cgcagccacaagattcaaaccagctcaactcatccaccgacaatcttccccctgtc 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1191 cgcagccacaagattcaaaaccagctcaactcatccaccgacaatcttccccctgtc 1250

Qy     1201 aacaagtttccgctgtaaagaagggggccccaccctcagttctgtgccatttact 1260
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1251 accagttccgctgtaagaaggggcgcccaccctcagttctgtgccatttact 1310

Qy     1261 tgcctcccaagtaccagctccgtcccggaggagggagcagatgccaattattact 1320
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1311 tgcctcccaagtaccagctccgtcccggaggagggagcagatgccaattattact 1370

Qy     1321 tgcaatcctgaggaattcatagttgagtgcggcagctttcccaagaaccggctg 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1371 tgcaatcctgaggaattcatagttgagtgcggcagcttccccaagcacgacgctg 1430

Qy     1381 caggagtacaggagtgctgcggcacggccccagtcaggagaagaaagtcagtac 1440
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1431 caggagtacaggagtgctgccggcacggccccagtcaggagaagaaagtcagtac 1490

Qy     1441 ccagaaatcatctcttcttggaacacagggtctgccatcccgatgaaagtcagt 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1491 ccagaaatcatctcttcttggaacacagggtctgccatcccgatgaaagtcagt 1550

Qy     1501 gcccacacttgtcaacataagcccgacacgtctctgctactggactgtggtcggcaca 1560
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1551 gccacacttgtcaacataagcccgacacgtctctgctactggactgtggtgagggcaca 1610

Qy     1561 tttgggcgctgtgccgtcattacgcagacaagtggcaggtcctgggccacctggct 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1611 tttgggcagctgtgccgtcattacggagacaagtggcaggtcctgggccacctggt 1670

Qy     1621 gctgtgttcgtgtcccacaatctcaccacacggctcaccaagtatcttgctg 1680
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1671 gctgtgttcgtgtcccacctgcaccagatcacacacggcttgccaagtatctgctg 1730

Qy     1681 cagagagaacgcgcctggcatctttgggaaacgccagtccctttgctgttgcc 1740
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1731 caggagagaacgcgcctcttgctcatctttgggaaacgccagtcccttgctggttgcc 1790

Qy     1741 cccaaacaagctcaaagctggtcttgctcttgcacaacctgccagggaggtcctgcac 1800
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1791 cccaaccagctcaaagctggtcttgctcttgcacaacctgccagggaggtcctgcac 1850

Qy     1801 cacatcagtatgattcctgccaaatgctcagaagggctgagatctccagtcctga 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1851 cacatcagtatgattcctgccaaatgcttcaggaaggggctgagatctccagtcctgca 1910
```

```
Qy  1861  gtggaaagattgatcagttctcgctgttccgaacctgtgattcggaaagagtttcagacctgt  1920
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  1911  gtggaaagattgatcagttctcgctgttccgaacatgtgattcggaaagagtttcagacctgt  1970

Qy  1921  ctggtgccggcactgcaagacatgccttggctctgtctgctggtgcacacctctggctggaaa  1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1971  ctggtgccggcactgcaagacatgccttggctctgtctgctggtgcacacctctggctggaaa  2030

Qy  1981  gtggtctattccgggacaccacatgccctcgcgaggctctggtcggatggggaaagatgcc  2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2031  gtggtctattccgggacaccacatgccctcgcgaggctctggtcggatggggaaagatgcc  2090

Qy  2041  accctcctgataacatgaagcacacctggaagatggtgttcggaaggaagcagtggaaag  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2091  accctcctgatacatgaagcacacctggaagatggtgttcggaaggaagcagtggaaag  2150

Qy  2101  acacacagcacacaacgtcccaagccatcagcgtggggatgaacgcggagttcatt  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2151  acacacagcacacaacgtcccaagccatcagcgtggggatgaacgcggagttcatt  2210

Qy  2161  atgtctgaaccaattcagccaccgcctatgccaaggtccccctcttcagcccaacttcagc  2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2211  atgtctgaaccaattcagccaccgcctatgccaaggtccccctcttcagcccaacttcagc  2270

Qy  2221  gagaaagtgggagttgccttgaccaacaatgaaggtctgcttggagactttccaacaatg  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2271  gagaaagtgggagttgccttgaccaacaatgaaggtctgcttggagactttccaacaatg  2330

Qy  2281  cccaagctgattccccactgttgttgcgcaatcgagccagttc  2340
          |||||||||||||||||||||||||||||||||||||||||||
Db  2331  cccaagctgattccccactgaaagccctgttgctgcgcaatcgagccagttc  2390

Qy  2341  cgcaggagagaacgggagctgcggcagtgcggcgcgcgccctccctgtcccagtcggca  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2391  cgcaggagagaacgggagctgcggcagtgcggcgcgcgccctccctgtcccagtcggca  2450

Qy  2401  gggcggctgaggatgggagctcagacagagaaggacgaggacacagttccacaggc  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2451  gggcggctgaggatgggagctcagacagagaaggacgaggacacagttccacaggc  2510

Qy  2461  aagaaggtcagagcccagtga  2481
          |||||||||||||||||||||
Db  2511  aagaaggtcagagcccagtga  2531
```

RESULT 2

AAA52810

ID   AAA52810 standard; cDNA; 2478 BP.
XX
AC   AAA52810;
XX
DT   20-SEP-2000 (first entry)
XX
DE   Human sulphatase G cDNA.
XX
KW   Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..2478
FT   /tag= a
FT   /partial
FT   /product= "hSG"
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999; 99WO-AU01092.
XX
PR   09-DEC-1998; 98AU-0007624.
XX
PN   WO200034327-A1.
XX
PT   Novel isolated DNA sequence which encodes human sulphatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency -
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ, Litjens T, Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   P-PSDB; AAY99850.
XX
CC   The present sequence encodes human sulphatase G (hSG). hSG is
CC   not a member of the well-characterised CTPSR sulphatase family.
CC   It belongs to a family showing sequence similarity to a sulphatase
CC   from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC   contains 23 exons and is located at chromosome 17p11.2. The present
CC   sequence is clone lambda29.1 from a human testes cDNA library. It was
CC   isolated using human EST sequences with sequence similarity to the
CC   non-CTPSR family as a probe to screen the library. The cDNA insert was
CC   subcloned and the DNA sequence of both strands was determined. The
CC   sequence may be used to treat a patient suffering from hSG deficiency
CC   by replacing, repairing, or compensating for a DNA sequence within that
CC   patient's genome.
XX
PS   Claim 2; Page 29-30; 33pp; English.
XX
SQ   Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;

Query Match   91.7%;   Score 2274;   DB 21;   Length 2478;
Best Local Similarity 99.8%;   Pred. No.0;
Matches 2474;   Conservative   0; Mismatches   4; Indels   0; Gaps   0;

```
Qy    1  atgtggcgcgcctttgctcgctgctcgcggtccggcgcgacgcacacatgtcgcaggacgc   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  atgtggcgcgcctttgctcgctgctcgcggtccggcgcgacgcacacatgtcgcaggacgc   60

Qy   61  accatatcgcaggcaccgccgcccgcgccgagcggcgcgcaaggagcccgctgcggcacctg  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  accatatcgcaggcaccgccgcccgcgccgagcggcgcgcaaggagcccgctgcggcacctg  120

Qy  121  cgcacgcgagagaacgcggaccgtcggggtgctccgggcgcccaaacaccgtgtacctg  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  cgcacgcgagagaacgcggaccgtcggggtgctccgggcgcccaaacaccgtgtacctg  180

Qy  181  caggtcgtgtggcaacgcgggtagccgggactcgggcgccgcgctcttctctccgagttc  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  ttaagtggaatgattccttacttacttaaaggaaaaccgggcttccaaagtgtgtacttctga  240

Qy  241  aaccggtatctcttcaactgtggagaagcggttcagagaactcatgcaggagcacaagtta  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  aaccggtatctcttcaactgtggagaagcggttcagagaactcatgcaggagcacaagtta  300

Qy  301  aaggttgctcggcctggacacatattcctgacacgaatgcactggtctaatgtggggggc  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  aaggttgctcggcctggacacatattcctgacacgaatgcactggtctaatgtggggggc  360

Qy  361  ttaagtggaatgattccttacttacttaaaggaaaaccgggcttccaaagtgtgtacttctga  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ttaagtggaatgattccttacttacttaaaggaaaaccgggcttccaaagtgtgtacttctga  420

Qy  421  cctccacaactgaaaataacctcgaagcaatcaaaaatatttctggtccattgaaagga  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  cctccacaactgaaaataacctcgaagcaatcaaaaatatttctggtccattgaaagga  480

Qy  481  atagaactgctgtcggggcccacctctgcccagaagatacgaggatgaaacatgacagtt  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  atagaactgctgtcggggcccacctctgcccagaagatacgaggatgaaacatgacagtt  540

Qy  541  taccagatcccaatacacacagtgaacaagagagggaaagcaccaaccaacatggcagagtcca  600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  taccagatcccaatacacacagtgaacaagagagggaaagcaccaaccaacatggcagagtcca  600
```

```
Qy    601  gaaaggcctctcagtcagctcaagtccagagaggcgaatcttcagagactcgaatgaaat  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  gaaaggcctctcagtcagctcaagtccagagagctcgagactcgaatgaaat  660

Qy    661  gagcacacacttcccaactctggttagctggccagtgactcttccctggtc  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  gagcacacacttcccaactctggttagctggccagtgactcttccctggtc  720

Qy    721  gtagctttcattcgttaagcttcaacttcttggtgctcaaagcaaag  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  gtagctttcactcgtaaagctcaacttcttggtgctcaaagcaaag  780

Qy    781  gagatgtggccctccagttgggacagctgccatccgctccgctccaatcttgtcaaggac  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  gagatgtggccctccagttgggacagctgccatccgctccgctccaatcttgtcaaggac  840

Qy    841  gggaaaagcacactcaatgaaggagaagagagatttggctgaagagctgttactcctcca  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  gggaaaagcacactcaatgaaggagaagagagatttggctgaagagctgttactcctcca  900

Qy    901  gatcctggtgctgcttttgtggtgttagaatgtccaga tgaaagcttcattcaaaccatc  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  gatcctggtgctgcttttgtggtgttagaatgtccaagcttcattcaaaccatc  960

Qy    961  tgtgagaatgccaacttcagaggtaccaaggcagatgccccgtggcctggtg  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  tgtgagaatgccaacttcagaggtaccaaggcagatgccccgtggcctggtg  1020

Qy   1021  gttcacatggcccagatctgtgcttgttgtggacagcagtacagtgtgatgagagg  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  gttcacatggcccagatctgtgcttgttgtggacagcagtacagtgtgatgagagg  1080

Qy   1081  tttgggcctaaagtaccagctccgtcccaggagggatggcagagagatgcca  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  tttgggcctaaagtaccagctccgtcccaggagggatggcagagagatgcca  1140

Qy   1141  cgcagccacagattcaatcaacctcatccaccggacaatcttccccctgctc  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  cgcagccacagattcaatcaacctcatccaccggacaatcttccccctgctc  1200

Qy   1201  accagttccgctgtaaggagaggccccacctcaagtgttccatggttcaggtgaa  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  accagttccgctgtaaggagaggccccacctcaagtgttccatggttcaggtgaa  1260

Qy   1261  tgcctcctcaagtaccagctccgtcccaggagggatggcagagagctgca  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  tgcctcctcaagtaccagctccgtcccaggagggatggcagagagctgca  1320

Qy   1321  tgcaatcctgaggaattcatagttgagggctgcagcttcccaacttcccagagcgtg  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  tgcaatcctgaggaattcatagttgagggctgcagcttcccaacttcccagagcgtg  1380

Qy   1381  caggagtacagaggagtgggcaggacggcccagcgcccacagaaaaagaagtcagtac  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  caggagtacagaggagtgggcaggacggcccagcgcccacagaaaaagaagtcagtac  1440

Qy   1441  ccagaaatcatcttccttgaacacgggtctgccatccgatgaagtgccagt  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  ccagaaatcatcttccttgaacacgggtctgccatccgatgaagtgccagt  1500

Qy   1501  gccacacttgtcaacataagcccgacacgtctctgctactggactgtgggcaca  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  gccacacttgtcaacataagcccgacacgtctctgctactggactgtgggcaca  1560

Qy   1561  tttgggcagctgtgccgtcattacggagaccaggtggacaggtcctggccacctggct  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  tttgggcagctgtgccgtcattacggagaccaggtggacaggtcctggccacctggct  1620

Qy   1621  gctgtgttgtgtccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg  1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  actgtgttgtgtccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg  1680

Qy   1681  cagagagaacgcgccttgcatctttgggaaagccgcttcaaccctttgctggttggc  1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  cagagagaacgcgccttgcatctttgggaaagccgcttcaaccctttgctggttggc  1740

Qy   1741  cccaaccagctcaaagctgctggccaagtaccaccagccaccagtgtccaggacctggtc  1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  cccaaccagctcaaagctgctggccaagtaccaccagccaccagtgtccaggacctggac  1800

Qy   1801  cacatcagtatgatctgccaaatgcttcaggaaggggctggaacagctgccactgtcctgca  1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  cacatcagtatgatctgccaaatgcttcaggaaggggctggaatccaagtcctgca  1860

Qy   1861  gtggaaagaattgatcagttcgctgctgtgcgaacatgtgattgatttggacctgt  1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  gtggaaagaattgatcagttcgctgctgtgcgaacatgtgattgatttggacctgt  1920

Qy   1921  ctgggtgcggcacctgcaagcagcatggttcgcgctgtgcacctctggaaa  1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  ctgggtgcggcacctgcaagcagcatggttcgcgctgtgcacctctggaaa  1980

Qy   1981  gtggtctattccgggacaccatgcccctgcgaagctcctggtccggatgggaaagatgcc  2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  gtggtctattccgggacaccatgcccctgcgaagctcctggtccggatgggaaagatgcc  2040

Qy   2041  accctcctgataacatgaagccacccctggaagtctcgcttgctgggctatggggaaat  2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  accctcctgataacatgaagccacccctggaagtctcgcttgctgggctatggggaaaat  2100

Qy   2101  acacacagcacacacgtccaagccatcagcgttgggatgcgggagttcatt  2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  acacacagcacacacgtccaagccatcagcgtggggatgcgggagttcatt  2160

Qy   2161  atgctgaaccacttcagcccagcgctatgccaagtccttcggatgggaacttcagc  2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161  atgctgaaccacttcagcccagcgctatgccaagtccttcggatgggaacttcagc  2220

Qy   2221  gagaaagtgaagagttgcctttgaccacaatgaaggtctgcttcttcccaacaatg  2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  gagaaagtgaagagttgcctttgccttggaagtctgcttcttcccaacaatg  2280

Qy   2281  cccaagctcgattccccactgaaagccccactgtttcgtcgaacatcgagggagtgaagag  2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  cccaagctcgattccccactgaaagccccactgtttcgtcgaacatcgagggagttgaggag  2340

Qy   2341  cgcaggggagaagcggagctgcggcggtcgcgggggccctcctgtccaggggagctgca  2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  cgcaggggagaagcggagctgcggcggtcgcgggggccctcctgtccaggggagctgca  2400

Qy   2401  gcgcgcctgaaggggagcctcagcagagcgccccacacagaggagccacaggcc  2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401  ggccggcctggaggatgggggagcctcagcagaagcgggccacacagaggagccacaggcc  2460

Qy   2461  aagaaggtcagagcccag  2478
           ||||||||||||||||||
Db   2461  aagaaggtcagagcccag  2478
```

RESULT   3

```
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; diabetes mellitus; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US008621.
XX
PR   31-MAR-1999; 99US-0127607.
PR   02-APR-1999; 99US-0127636.
PR   05-APR-1999; 99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42236.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3179-3180; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;

Query Match             64.5%;  Score 1600;  DB 21;  Length 2546;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1720;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;
```

```
Qy    762 cttggtgtctcaaagcaaaggaggggcctccacgctgccatgtccccat 811
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 cttggtgtctcaaagcaaaggagatggggctccacgctggccatgtccccat 419

Qy    822 cattgctgctgtcaaggacgggaaagctcatcactcaatgaaggaagaatttggtctga 881
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 cattgctgctgtcaaggacgggaaagctcatcactcaatgaaggaagaatttggctga 479

Qy    882 agagctgtgtactcctccagatacccctgctgcttgttggtgtggagaatgtccagatga 941
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 agagctgtgtactcctccagatacccctgctgcttgttggtgtggagaatgtccagatga 539

Qy    942 aagcttcattcaaccacatcttgtgagaaatgcatccttcagaggttctcctgtggtgaca 1001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 aagcttcattcaaccacatcttgtgagaaatgcatccttcagaggttctcctgtggtgaca 599

Qy   1002 tgcccccgtggccttggtgtggttcacatggccccagccgatctgtgttgttgctca 1061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 tgcccccgtggccttggtgtggttcacatggccccagccgatctgtgttgttgctca 659

Qy   1062 ccagcagtgatggagagggtttgggcctgaacaccagcactgtt 1121
          ||||||||||||||||||||||||||||||||||||||||||||
Db    660 ccagcagtgatggagagggtttgggcctgaacaccagcactgtt 719

Qy   1122 tgcctcagttcacacaccttcgcagccacaagattcaaaccca 1181
          ||||||||||||||||||||||||||||||||||||||||||||
Db    720 tgcctcagttcacacaccttcgcagccacaagattcaaaccca 779

Qy   1182 ggacatcttcccctgctcaccagtttccgctgtaagagggccc 1241
          ||||||||||||||||||||||||||||||||||||||||||||
Db    780 ggacatcttcccctgctcaccagtttccgctgtaagagggccc 839

Qy   1242 gcccatgttcaggtgaatgcctcctcaagtaccagcactccgt 1301
          ||||||||||||||||||||||||||||||||||||||||||||
Db    840 gcccatgttcaggtgaatgcctcctcaagtaccagcactccgt 899

Qy   1302 gagggatgcattattacttgcaatcctgaggaattcata 1361
          ||||||||||||||||||||||||||||||||||||||||
Db    900 gagggatgcattattacttgcaatcctgaggaattcata 959

Qy   1362 caacttccagacgacgtgcaggagtacagaggagtgcg 1421
          ||||||||||||||||||||||||||||||||||||||
Db    960 caacttccagacgacgtgcaggagtacagaggagtgcg 1019

Qy   1422 agagaaagaagtcagtacccagaaatcatcctt 1481
          |||||||||||||||||||||||||||||||||
Db   1020 agagaaagaagtcagtacccagaaatcatcctt 1079

Qy   1482 gaagattcgaaatgtcagtgccacaactgtc 1541
          |||||||||||||||||||||||||||||||
Db   1080 gaagattcgaaatgtcagtgccacaactgtc 1139

Qy   1542 ggactgtggtgagggcacatttggcaagctg 1601
          |||||||||||||||||||||||||||||||
Db   1140 ggactgtggtgagggcacatttggcaagctg 1199

Qy   1602 ggtcctgggcaccctgctgcgtgttgttgtc 1661
          |||||||||||||||||||||||||||||||
Db   1200 ggtcctgggcaccctgctgcgtgttgttgtc 1259

Qy   1662 cttgccaagtatcttgctgcaagacacgcgc 1721
          |||||||||||||||||||||||||||||||
Db   1260 cttgccaagtatcttgctgcaagacacgcgc 1319

Qy   1722 cccttgcttggttggttgccccaccacacct 1781
          |||||||||||||||||||||||||||||||
Db   1320 cccttgcttggttggttgccccaccacacct 1379

Qy   1782 gtgcaggaggtcctgcacacatcagtat 1841
          ||||||||||||||||||||||||||||
Db   1380 gtgcaggaggtcctgcacacatcagtat 1439

Qy   1842 tgagatctcagtcctgcagtggaagaatt 1901
          |||||||||||||||||||||||||||||
Db   1440 tgagatctcagtcctgcagtggaagaatt 1499

Qy   1902 ggaaagagtttcaagacttcgttctgg 1961
          |||||||||||||||||||||||||||
Db   1500 ggaaagagtttcaagacttcgttctgg 1559

Qy   1962 gcacacctcggctggaaagtgttcatt 2021
          |||||||||||||||||||||||||||
```

```
Db   1560 gcacctctgctgaaagtggtctattccgggggaccaatgccctgcggagctctggt 1619
Qy   2022 ccggatgggaaagatgccacctctcgatacatgaagcaccctggaaagatggtttgga 2081
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 ccggatgggaaagatgccacctctcgatacatgaagcaccctggaaagatggtttgga 1679
Qy   2082 agaggaagcagtggaaaagacacacaacagccacatcaagctgggggatgcg 2141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 agaggaagcagtggaaaagacacacaacagccacatcaagctgggggatcg 1739
Qy   2142 gatgaacgcggagttcattatgctgaaccacctgcgtatccgcaaggtcccct 2201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 gatgaacgcggagttcattatgctgaaccacctgcgtatccgcaaggtccccct 1799
Qy   2202 cttcagcccccaacttcagcgagaaagtgggagttgcctttgaccacatgaaggtctgctt 2261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 cttcagcccccaacttcagcgagaaagtgggagttgcctttgaccacatgaaggtctgctt 1859
Qy   2262 tggagactttccaacaatgccaagctgatt-ccccactgaaagccctgttttgctggcg 2320
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1860 tggagactttccaacaatgccaagctgattccccactgaaagccctgttttgctggcg 1919
Qy   2321 acatcgaggagatgggagagacgcagatcgcgggcaggtgcggcggggcc 2380
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 acatcgaggagatgggagagacgcagatcgcggggcaggtgcggcggggccc 1979
Qy   2381 tcctgtccaggagagctgccagggggctcggagaggaagcctcagcagaagcgggccc 2440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1980 tcctgtccaggagagctgccaggggatgggggagcctcagcagaagcgggccc 2039
Qy   2441 acacagaggagccacaggccaagaaggtcagagcccagtga 2481
          ||||||||||||||||||||||||||||||||||||||||
Db   2040 acacagaggagccacaggccaagaaggtcagagcccagtga 2080
```

```
FH  Key          Location/Qualifiers
FH
FT  CDS          910..26039
FT               /*tag= a
FT               /product= "HPC2"
FT               /note= "this sequence contains introns"
FT               /transl_except= (pos:23892..23895,aa:Glu)
FT  exon         910..1154
FT               /*tag= b
FT               /number= 1
FT  exon         1736..1786
FT               /*tag= c
FT               /number= 2
FT  exon         1925..1995
FT               /*tag= d
FT               /number= 3
FT  exon         3025..3089
FT               /*tag= e
FT               /number= 4
FT  exon         4361..4418
FT               /*tag= f
FT               /number= 5
FT  exon         5582..5650
FT               /*tag= g
FT               /number= 6
FT  exon         7075..7194
FT               /*tag= h
FT               /number= 7
FT  exon         8186..8244
FT               /*tag= i
FT               /number= 8
FT  exon         12878..12936
FT               /*tag= j
FT               /number= 9
FT  exon         13032..13104
FT               /*tag= k
FT               /number= 10
FT  exon         13756..13868
FT               /*tag= l
FT               /number= 11
FT  exon         15283..15378
FT               /*tag= m
FT               /number= 12
FT  exon         16278..16416
FT               /*tag= n
FT               /number= 13
FT  exon         16498..16583
FT               /*tag= o
FT               /number= 14
FT  exon         18583..18701
FT               /*tag= p
FT               /number= 15
FT  exon         20349..20445
FT               /*tag= q
FT               /number= 16
FT  exon         22172..22310
FT               /*tag= r
FT               /number= 17
FT  exon         22879..22917
FT               /*tag= s
FT               /number= 18
FT  exon         23045..23154
FT               /*tag= t
FT               /number= 19
FT  exon         23795..23895
FT               /*tag= u
FT               /number= 20
FT  exon         23973..24093
FT               /*tag= v
FT               /number= 21
FT  exon         24354..24432
FT               /*tag= w
FT               /number= 22
FT  exon         25026..25170
FT               /*tag= x
FT               /number= 23
FT  exon         25812..26036
FT               /*tag= y
FT               /number= 24
FT  polyA_signal 26447..26452
FT               /*tag= z
```

PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Claim 3; Page 108-122; 157pp; English.
XX
CC   The present sequence is the genomic sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;

Query Match   100.0%; Score 247; DB 21; Length 26664;
Best Local Similarity   100.0%; Pred. No. 5.4e-111;
Matches 247; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 atgtgggcgcttctgctcgctcgcggtccgcggccgagacgcaccacatgtcgcaggagc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   910 atgtgggcgcttctgctcgctcgcggtccgcggccgagacgcaccacatgtcgcaggagc 969

Qy   61 accatatcgcaggcaccgcccgcccgcccgagcgagccgcgcaccgaagagaccgcgcaactg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   970 accatatcgcaggcaccgcccgcccgcccgagcgagccgcgcaccgaagagaccgcgcaactg 1029

Qy   121 cgcacgcgagagaagcccggaaccgtcggggtgtctccggcgcgccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1030 cgcacgcgagagaagcccggaccgtcggggtgtctccggcgcgccaaacaccgtgtacctg 1089

Qy   181 caggtggtggcagcgggtagccgggactccggagcgccgcgcgctctacgtctttccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1090 caggtggtggcagcgggtagccgggactccggagcgccgcgcgctctacgtctttccgagttc 1149

Qy   241 aaccggt 247
          |||||||
Db   1150 aaccggt 1156

RESULT 5
AA280231
ID   AA280231 standard; cDNA; 238 BP.
XX
AC   AA280231;
XX
DT   07-APR-2000 (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW   hyperplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO9964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999; 99WO-IB01062.
XX
PR   10-JUN-1998; 98US-0088801.
XX
PA   (FARB) BAYER CORP.
XX
PI   Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI   Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI   Schlegel R;

XX
DR   WPI; 2000-087220/07.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer -
XX
PS   Claim 15; Page 258; 469pp; English.
XX
CC   AA279917 to AA280766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which
CC   can be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g.
CC   carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC   can be used for developing agents for the diagnosis and treatment of
CC   disorders involving unwanted cell proliferation, such as neoplasia,
CC   dysplasia or hyperplasia.
XX
SQ   Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;

Query Match   9.6%; Score 237; DB 21; Length 238;
Best Local Similarity   100.0%; Pred. No. 4.8e-106;
Matches 237; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   176 acctgcaagtggttgcagcgggggttagccgggactccgggcgccgcgctctacgtctttccg 235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 acctgcaagtggttgcagcgggggttagccgggactccgggcgccgcgctctacgtctttccg 60

Qy   236 agttcaacccgtatctcttcaacctgtggagaagcggttcagagagactcatgcaggagcaca 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 agttcaacccgtatctcttcaacctgtggagaagcggttcagagagactcatgcaggagcaca 120

Qy   296 agttaaaggttgctccgcctggacacacatattccctggacacgaatgtcactgttctaatgttg 355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 agttaaaggttgctccgcctggacacacatattccctggacacgaatgtcactgttctaatgttg 180

Qy   356 ggggcttaagtggaatgattccttacttaaaggaaccgggcttccaaagtgtgtac 412
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ggggcttaagtggaatgattccttacttaaaggaaccgggcttccaaagtgtgtac 237

RESULT 6
AAA60390
ID   AAA60390 standard; cDNA; 326 BP.
XX
AC   AAA60390;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Murine prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Mouse; prostate cancer predisposing gene; HPC2; gene therapy;
KW   peptide therapy; drug design; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FT   CDS   51..269
FT   /*tag= a
FT   /product= "Mm.HPC2"
FT   /partial
FT   /note="this sequence contains no termination codon"
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.

```
XX
PA  (MYRI-) MYRIAD GENETICS INC.
XX
PI  Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR  WPI; 2000-376481/32.
DR  P-PSDB; AAB07230.
XX
PT  Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT  antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS  Disclosure; Page 151; 157pp; English.
XX
CC  The present sequence is the coding sequence of the murine prostate
CC  cancer predisposing gene Mn. HPC2, the human homologue of which is found
CC  on chromosome 17p. Some alleles of this gene cause a predisposition to
CC  cancer, particularly prostate cancer. This gene and its protein can be
CC  used in peptide and gene therapy for cancer patients, as well as being
CC  useful as diagnostic tools (both for cancer sufferers and those with a
CC  predisposition to the disease) and in the production of cancer drugs.
XX
SQ  Sequence 326 BP; 39 A; 104 C; 127 G; 56 T; 0 other;

Query Match             1.0%;  Score 26;  DB 21;  Length 326;
Best Local Similarity   100.0%;  Pred. No. 0.01;
Matches  26;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  166 aacaccgtgtacctgcaggtggtggc 191
        ||||||||||||||||||||||||||
Db  192 aacaccgtgtacctgcaggtggtggc 217


RESULT  7
AAA60373
ID  AAA60373 standard; cDNA; 23 BP.
XX
AC  AAA60373;
XX
DT  07-DEC-2000  (first entry)
XX
DE  Human prostate cancer predisposing gene HPC2 variant 1641insG (normal).
XX
KW  Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW  gene therapy; peptide therapy; drug design; ss.
XX
OS  Homo sapiens.
XX
PN  WO200027864-A1.
XX
PD  18-MAY-2000.
XX
PF  05-NOV-1999;  99WO-US26055.
XX
PR  06-NOV-1998;  98US-0107468.
XX
PA  (MYRI-) MYRIAD GENETICS INC.
XX
PI  Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR  WPI; 2000-376481/32.
XX
PT  Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT  antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS  Example 5; Page 63; 157pp; English.
XX
CC  The present sequence is part of the coding sequence of a variant of the
CC  human prostate cancer predisposing gene HPC2, which is found on
CC  chromosome 17p. This variant has been designated 1641insG. Some
CC  alleles of this gene cause a predisposition to cancer, particularly
CC  prostate cancer, this allele being an example of this. The HPC2 gene and
CC  its protein can be used in peptide and gene therapy for cancer patients,

CC  as well as being useful as diagnostic tools (both for cancer sufferers
CC  and those with a predisposition to the disease) and in the production of
CC  cancer drugs. This sequence was isolated by mutation screening of the
CC  HPC2 gene in humans.
XX
SQ  Sequence 23 BP; 5 A; 9 C; 5 G; 4 T; 0 other;

Query Match             0.9%;  Score 23;  DB 21;  Length 23;
Best Local Similarity   100.0%;  Pred. No. 0.32;
Matches  23;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  1630 gtgtcccacctgcacgcagatca 1652
         |||||||||||||||||||||||
Db     1 gtgtcccacctgcacgcagatca 23


RESULT  8
AAC24564/c
ID  AAC24564 standard; cDNA; 143 BP.
XX
AC  AAC24564;
XX
DT  06-OCT-2000  (first entry)
XX
DE  Human secreted protein 5' EST, SEQ ID NO: 28639.
XX
KW  Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW  gene therapy; chromosome mapping; ss.
XX
OS  Homo sapiens.
XX
PN  EP1033401-A2.
XX
PD  06-SEP-2000.
XX
PF  21-FEB-2000;  2000EP-0200610.
XX
PR  26-FEB-1999;  99US-0122487.
XX
PA  (GEST ) GENSET.
XX
PI  Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR  WPI; 2000-500381/45.
XX
PT  New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT  obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT  diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS  Claim 1; SEQ ID 28639; 71pp + CD-ROM; English.
XX
CC  The present sequence is one of a large number of 5' ESTs derived from
CC  mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC  identified within the present sequence. The 5' ESTs were prepared from
CC  total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC  sequences usually correspond mainly to the 3' untranslated region (UTR)
CC  of the mRNA because they are often obtained from oligo-dT primed cDNA
CC  libraries. Such ESTs are not well suited for isolating cDNA sequences
CC  derived from the 5' ends of mRNAs and even in those cases where longer
CC  cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC  5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC  used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC  in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC  They are used to obtain upstream regulatory sequences and to design
CC  expression and secretion vectors.
XX
SQ  Sequence 143 BP; 40 A; 26 C; 45 G; 32 T; 0 other;

Query Match             0.8%;  Score 21;  DB 21;  Length 143;
Best Local Similarity   100.0%;  Pred. No. 2.9;
Matches  21;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

Qy   976 tttcagaggttaccaaggaaag 996
          ||||||||||||||||||||||
Db   67 TTTCAGAGGTACCAAGGAAAG 47

RESULT 9
ID   AAA60275/c
XX
AC   AAA60275;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA exon 1 mutation screening primer SEQ ID NO: 96.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer   -
XX
PS   Example 5; Page 59; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers, and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 38 BP; 10 A; 9 C; 11 G; 8 T; 0 other;

Query Match   0.8%; Score 20; DB 21; Length 38;
Best Local Similarity   100.0%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   56 gacgcaccatatcgcaggca 75
          ||||||||||||||||||||
Db   38 GACGCACCATATCCGCAGGCA 19

RESULT 10
ID   AAF09297
XX
AC   AAF09297;
XX
DT   13-MAR-2001 (first entry)
XX
DE   Fusarium venenatum EST SEQ ID NO:1820.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;

KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Fusarium venenatum.
XX
PN   WO20005762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999; 99US-0273623.
XX
PA   (NOVO) NOVO NORDISK BIOTECH INC.
PA   (NOVO) NOVO NORDISK AS.
XX
PI   Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags   -
XX
PS   Claim 86; Page 1064; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 308 BP; 88 A; 56 C; 100 G; 63 T; 1 other;

Query Match   0.8%; Score 20; DB 21; Length 308;
Best Local Similarity   100.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   2334 ggaggagcgccaggagaagc 2353
          ||||||||||||||||||||
Db   30 ggaggagcgccaggagaagc 49

RESULT 11
ID   AAC57035
XX
AC   AAC57035;
XX
DT   25-JAN-2001 (first entry)

DE   Pinus radiata transcription factor DNA sequence #481.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Pinus radiata.
XX
PN   WO20005372A-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999; 99US-0266513.
PR   18-AUG-1999; 99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
-
PS   Claim 1; Pages 570-571; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 457 BP; 130 A; 94 C; 118 G; 115 T; 0 other;

Qy   2082 agaggaagcagtggaaaga 2101
             ||||||||||||||||||||
Db   146 agaggaagcagtggaaaga 165

Query Match   0.8%; Score 20; DB 21; Length 457;
Best Local Similarity   100.0%;   Pred. No. 8.6;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 12
AAA60211/c
ID   AAA60211 standard; DNA; 19 BP.
XX
AC   AAA60211;
XX
DT   07-DEC-2000   (first entry)
XX
DE   Human HPC2 cDNA 5' RACE primer SEQ ID NO: 32.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.

XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 19 BP; 4 A; 7 C; 3 G; 5 T; 0 other;

Qy   257 actgtgggagaaggcgttca 275
             ||||||||||||||||||||
Db   19 ACTGTGGGAGAAGGCGTTCA 1

Query Match   0.8%; Score 19; DB 21; Length 19;
Best Local Similarity   100.0%;   Pred. No. 29;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 13
AAA60229
ID   AAA60229 standard; DNA; 19 BP.
XX
AC   AAA60229;
XX
DT   07-DEC-2000   (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 50.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.

XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
PT

```
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
CC
SQ   Sequence 19 BP; 5 A; 3 C; 6 G; 5 T; 0 other;
XX

Query Match    Best Local Similarity  100.0%;  Score 19;  DB 21;  Length 19;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1901 tggaagagtttcagacctg 1919
           |||||||||||||||||||
Db       1 tggaagagtttcagacctg 19

RESULT 14
AAA60230/c
ID   AAA60230 standard; DNA; 19 BP.
XX
AC   AAA60230;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 51.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO20002787864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
XX
PS   Example 3; Page 55; 157pp; English.

The present sequence is a primer used in the isolation of the human
and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
version of the gene is found on chromosome 17p. Some alleles cause a
predisposition to cancer, particularly prostate cancer. This gene and its
protein can be used in peptide and gene therapy for cancer patients, as
well as being useful as diagnostic tools (both for cancer sufferers and
those with a predisposition to the disease) and in the production of
cancer drugs.

Sequence 19 BP; 5 A; 6 C; 3 G; 5 T; 0 other;

Query Match    Best Local Similarity  100.0%;  Score 19;  DB 21;  Length 19;
```

```
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1901 tggaagagtttcagacctg 1919
           |||||||||||||||||||
Db      19 TGGAAGAGTTTCAGACCTG 1

RESULT 15
AAA60233
ID   AAA60233 standard; DNA; 19 BP.
XX
AC   AAA60233;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 54.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO20002787864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 19 BP; 8 A; 5 C; 3 G; 3 T; 0 other;
XX

Query Match    Best Local Similarity  100.0%;  Score 19;  DB 21;  Length 19;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    427 caactggaaaatacctcg 445
          ||||||||||||||||||
Db      1 caactggaaaatacctcg 19

RESULT 16
AAA60235
ID   AAA60235 standard; DNA; 19 BP.
XX
AC   AAA60235;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 56.
XX
```

KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 19 BP; 5 A; 3 C; 6 G; 5 T; 0 other;

Query Match   0.8%;   Score 19;   DB 21;   Length 19;
Best Local Similarity   100.0%;   Pred. No. 29;
Matches   19;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   751 agaggaaacttcttggtgc 769
         |||||||||||||||||||
Db   1 agaggaaacttcttggtgc 19

RESULT 17
AAA60364
ID   AAA60364 standard; DNA; 19 BP.
XX
AC   AAA60364;
XX
DT   07-DEC-2000   (first entry)
XX
DE   Human HPC2 cDNA exon 24 3'UTR mutation screening primer SEQ ID NO: 185.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;

XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 5; Page 62; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 19 BP; 7 A; 5 C; 5 G; 0 U; 0 other;

Query Match   0.8%;   Score 19;   DB 21;   Length 19;
Best Local Similarity   100.0%;   Pred. No. 29;
Matches   19;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy 2439 ccacacagaggagccacag 2457
         |||||||||||||||||||
Db   1 ccacacagaggagccacag 19

RESULT 18
AAA60214
AAA60214/c
ID   AAA60214 standard; DNA; 29 BP.
XX
AC   AAA60214;
XX
DT   07-DEC-2000   (first entry)
XX
DE   Human HPC2 cDNA expression construct primer SEQ ID NO: 35.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.

SQ   Sequence 29 BP; 4 A; 10 C; 7 G; 8 T; 0 other;

Qy   2463 gaaggtcagagccaagtga 2481
          |||||||||||||||||||
Db   29 GAAGGTCAGAGCCAGTGA 11

Query Match   0.8%; Score 19; DB 21; Length 29;
Best Local Similarity   100.0%; Pred. No. 29;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 19
ID   AAA60362 standard; DNA; 37 BP.
XX
AC   AAA60362;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA exon 24 ORF mutation screening primer SEQ ID NO: 183.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
PT   06-NOV-1998; 98US-0107468.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
PS   Example 5; Page 62; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 37 BP; 10 A; 13 C; 9 G; 5 T; 0 other;

Qy   2439 ccacacagagagccacag 2457
          ||||||||||||||||||
Db   19 ccacacagagagccacag 37

Query Match   0.8%; Score 19; DB 21; Length 37;
Best Local Similarity   100.0%; Pred. No. 28;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 20
ID   AAA60307 standard; DNA; 38 BP.
XX
AC   AAA60307;

XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA exon 9 mutation screening primer SEQ ID NO: 128.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer
XX
PS   Example 5; Page 60; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 38 BP; 14 A; 6 C; 12 G; 6 T; 0 other;

Qy   810 catcgctcccaccaattgct 828
          ||||||||||||||||||||
Db   38 CATCGCTCCCATCATTGCT 20

Query Match   0.8%; Score 19; DB 21; Length 38;
Best Local Similarity   100.0%; Pred. No. 28;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 21
ID   AAV88751/c
ID   AAV88751 standard; cDNA; 526 BP.
XX
AC   AAV88751;
XX
DT   12-FEB-1999 (first entry)
XX
DE   EST clone HK189.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845437-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998; 98WO-US06956.

XX
PR   10-APR-1997;   97US-0837312.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI   Racie LA, Spaulding V, Treacy M;
XX
DR   WPI; 1999-070078/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 500; 641pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoieisis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 526 BP; 127 A; 138 C; 136 G; 125 T; 0 other;

Query Match   0.8%;   Score 19;   DB 20;   Length 526;
Best Local Similarity   100.0%;   Pred. No. 26;
Matches 19;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   692 gaagaggggtcaggggactc 710
          ||||||||||||||||||||
Db   272 GAAGAGGGGTCAGGGACTC 254

RESULT 22
ID   AAV82020 standard; DNA; 7642 BP.
XX
AC   AAV82020;
XX
DT   21-JUN-1999 (first entry)
XX
DE   Moraxella catarrhalis Q8 lfr region.
XX
KW   Lactoferrin receptor; lactoferrin binding protein; LBP1; LBP2;
KW   lbpA gene; lbpB gene; ORF3; infection; otitis media; sinusitis;
KW   conjunctivitis; pneumonia; bronchitis; tracheitis; emphysema;
KW   diagnosis; therapy; vaccine; Branhamella catarrhalis; ss.
XX
OS   Moraxella catarrhalis.
XX
FH   Key   Location/Qualifiers
FH
FT   -35_signal   90..95
FT   /*tag= a
FT   -10_signal   108..113
FT   /*tag= b
FT   RBS   138..147
FT   /*tag= c
FT   CDS   152..287
FT   /*tag= d
FT   /gene= lbpB
FT   /product= LBP2
FT   /transl_except= (pos:509..512, aa:Asp)
FT   /note= "this codon has an apparent 1 codon

FT   insertion, which alters the reading frame"
FT   /note= "lbpB is specifically claimed in Claim 8;
FT   encodes AAW89417"
FT   -35_signal   2974..2979
FT   /*tag= e
FT   -10_signal   2991..2996
FT   /*tag= f
FT   RBS   3007..3016
FT   /*tag= g
FT   CDS   3022..6024
FT   /*tag= h
FT   /gene= lbpA
FT   /product= LBP1
FT   /note= "specifically claimed in Claim 8; encodes
FT   AAW89418"
FT   CDS   3067..6024
FT   /*tag= i
FT   /gene= lbpA
FT   /product= LBP1
FT   /note= "specifically claimed in claim 8; encodes
FT   AAW89419"
FT   CDS   6026..7642
FT   /*tag= j
FT   /gene= orf3
FT   /note= "specifically claimed in Claim 7; encodes
FT   AAW89419"
XX
PN   WO9855606-A2.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-CA00544.
XX
PR   08-MAY-1998;   98US-0074658.
PR   03-JUN-1997;   97US-0867941.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Du R, Klein MH, Loosmore SM, Wang Q, Yang Y;
XX
DR   WPI; 1999-070266/06.
DR   P-PSDB; AAW89417, AAW89418, AAW89419, AAW89420.
XX
PT   Lactoferrin receptor genes from Moraxella, especially M. catarrhalis
PT   - useful to diagnose Moraxella infection e.g. to detect otitis media
PT   due to M. catarrhalis infection and to immunise against such
PT   infections
XX
PS   Claim 8; Fig 4A-4P'; 202pp; English.
XX
CC   This polynucleotide comprises the lactoferrin receptor (lfr) locus
CC   of Moraxella catarrhalis (Branhamella catarrhalis) Q8. There are
CC   3 tandem genes in locus, identified as lbpA, lbpB (alternative
CC   start codons) and orf3A, respectively encoding lactoferrin binding
CC   protein 2 (Lbp2, see AAW89417), lactoferrin binding protein 1 (Lbp1,
CC   see AAW89418 and AAW89419) and open reading frame protein 3 (ORF3, see
CC   AAW89420). The lfr locus was identified following generation of a M.
CC   catarrhalis strain Q8 genomic DNA library and screening with
CC   specific hybridisation probes. The genes and DNA sequences of the
CC   lfr locus are useful for diagnosis, immunisation, and the
CC   generation of diagnostic and immunological reagents. Immunogenic
CC   compositions, including vaccines, based upon expressed recombinant
CC   Lbp1 and/or Lbp2 and/or ORF3, portions of these or their analogues,
CC   can be prepared for prevention of diseases caused by Moraxella. M.
CC   catarrhalis is a causative agent of otitis media and has been
CC   associated with sinusitis, conjunctivitis and inflammatory diseases
CC   of the lower respiratory tract, such as pneumonia, chronic
CC   bronchitis, tracheitis and emphysema.
XX
SQ   Sequence 7642 BP; 2417 A; 1726 C; 1631 G; 1868 T; 0 other;

Query Match   0.8%;   Score 19;   DB 20;   Length 7642;

```
Best Local Similarity, 100.0%; 0; Pred. No. 25; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0;

Qy    761 tcttggtgtcaaagcaaa 779
          ||||||||||||||||||
Db   3657 tcttggtgtcaaagcaaa 3675
```

RESULT 23

AAA60219
ID   AAA60219 standard; DNA; 18 BP.
XX
AC   AAA60219;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 40.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786 4-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 3 A; 2 C; 6 G; 7 T; 0 other;

```
Query Match    0.7%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    346 tctaatgtggggggctta 363
          ||||||||||||||||||
Db      1 tctaatgtggggggctta 18
```

RESULT 24

AAA60220/c
ID   AAA60220 standard; DNA; 18 BP.
XX
AC   AAA60220;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 41.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786 4-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 7 A; 6 C; 2 G; 3 T; 0 other;

```
Query Match    0.7%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    346 tctaatgtggggggctta 363
          ||||||||||||||||||
Db     18 TCTAATGTGGGGGCTTA 1
```

RESULT 25

AAA60221
ID   AAA60221 standard; DNA; 18 BP.
XX
AC   AAA60221;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 42.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786 4-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 7 A; 6 C; 2 G; 3 T; 0 other;

```
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
     antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 7 A; 5 C; 3 G; 3 T; 0 other;

Qy   654 tgaaaatgagccacact 671
         ||||||||||||||||||
Db     1 tgaaaatgagccacact 18

Query Match             0.7%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT  26
AAA60222/c
ID   AAA60222 standard; DNA; 18 BP.
XX
AC   AAA60222;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 43.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002764-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
     antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
```

```
XX
SQ   Sequence 18 BP; 3 A; 3 C; 5 G; 7 T; 0 other;

Qy   654 tgaaaatgagccacact 671
         ||||||||||||||||||
Db    18 TGAAAATGAGCCACACCT 1

Query Match             0.7%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT  27
AAA60223
ID   AAA60223 standard; DNA; 18 BP.
XX
AC   AAA60223;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 44.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002764-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
     antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 6 C; 2 G; 5 T; 0 other;

Qy   948 cattcaaccatctgtga 965
         |||||||||||||||||||
Db     1 cattcaaccatctgtga 18

Query Match             0.7%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT  28
AAA60224/c
ID   AAA60224 standard; DNA; 18 BP.
XX
```

```
AC   AAA60224;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 45.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 2 C; 6 G; 5 T; 0 other;
XX
XX

Query Match   0.7%;   Score 18;   DB 21;   Length 18;
Best Local Similarity   100.0%;   Pred. No. 89;
Matches   18;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy  948 cattcaaccatctgtga 965
        |||||||||||||||||||
Db   18 CATTCAACCATCTGTGA 1


RESULT 29
AAA60225
ID   AAA60225 standard; DNA; 18 BP.
AAA60225;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 46.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
```

```
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer.
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 5 C; 3 G; 5 T; 0 other;
XX
XX

Query Match   0.7%;   Score 18;   DB 21;   Length 18;
Best Local Similarity   100.0%;   Pred. No. 89;
Matches   18;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy 1257 tgaatgcctcctcaagta 1274
        ||||||||||||||||||
Db    1 tgaatgcctcctcaagta 18


RESULT 30
AAA60226/c
ID   AAA60226 standard; DNA; 18 BP.
AC   AAA60226;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 47.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
```

```
CC      predisposition to cancer, particularly prostate cancer. This gene and its
CC      protein can be used in peptide and gene therapy for cancer patients, as
CC      well as being useful as diagnostic tools (both for cancer sufferers and
CC      those with a predisposition to the disease) and in the production of
CC      cancer drugs.
XX
SQ      Sequence 18 BP; 5 A; 3 C; 5 G; 5 T; 0 other;

Db      18 TGAATGCCTCCTCAAGTA 1
           ||||||||||||||||||
Qy    1257 tgaatgcctcctcaagta 1274

Matches    18; Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
Best Local Similarity  100.0%;  Pred. No. 89;
Query Match     0.7%;   Score 18;  DB 21;  Length 18;


RESULT 31
ID   AAA60227 standard; DNA; 18 BP.
XX
AC   AAA60227;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 48.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; HPC2;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 3 A; 3 C; 7 G; 5 T; 0 other;

Db   18 TGAATGCCTCCTCAAGTA 1
        ||||||||||||||||||
Qy 1257 tgaatgcctcctcaagta 1274

Query Match     0.7%;   Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches   18; Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy 1536 gctactggactgtggtga 1553
        ||||||||||||||||||
Db    1 gctactggactgtggtga 18
```

```
RESULT 32
ID   AAA60228/c
     AAA60228 standard; DNA; 18 BP.
XX
AC   AAA60228;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 49.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; HPC2;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 7 C; 3 G; 3 T; 0 other;

Db   18 GCTACTGGACTGGTGTGA 1
        ||||||||||||||||||
Qy 1536 gctactggactgggtga 1553

Query Match     0.7%;   Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 89;
Matches   18; Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 33
ID   AAA60232/c
     AAA60232 standard; DNA; 18 BP.
XX
AC   AAA60232;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 53.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; HPC2;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
```

```
RESULT 34
AAA60236
ID   AAA60236 standard; DNA; 18 BP.
XX
AC   AAA60236;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 57.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
     antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 6 A; 2 C; 7 G; 3 T; 0 other;

Qy   227 tctctccgagttcaacc 244
         ||||||||||||||||||
Db    18 TCTTCTCCGAGTTCAACC 1

Query Match             0.7%;   Score 18;   DB 21;   Length 18;
Best Local Similarity   100.0%;   Pred. No. 89;
Matches   18;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
```

```
RESULT 35
AAA60237
ID   AAA60237 standard; DNA; 18 BP.
XX
AC   AAA60237;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 58.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
     antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 8 A; 3 C; 6 G; 1 T; 0 other;

Qy   986 accaaggaaaggcagatg 1003
         ||||||||||||||||||
Db     1 accaaggaaaggcagatg 18

Query Match             0.7%;   Score 18;   DB 21;   Length 18;
Best Local Similarity   100.0%;   Pred. No. 89;
Matches   18;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
```

Qy   1510 gtcaacataagccccgac 1527
          ||||||||||||||||||
Db   1 gtcaacataagccccgac 18

RESULT 36
ID   AAA60238 standard; DNA; 18 BP.
XX
AC   AAA60238;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 59.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786A-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 0 A; 3 C; 7 G; 8 T; 0 other;

Query Match   0.7%; Score 18; DB 21; Length 18;
Best Local Similarity   100.0%; Pred. No. 89;
Matches   18; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1617 ggctgctgtgtttgtc 1634
          ||||||||||||||||||
Db   1 ggctgctgtgtttgtc 18

RESULT 37
ID   AAA60240 standard; DNA; 18 BP.
XX
AC   AAA60240;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 61.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;

KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786A-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 4 C; 3 G; 6 T; 0 other;

Query Match   0.7%; Score 18; DB 21; Length 18;
Best Local Similarity   100.0%; Pred. No. 89;
Matches   18; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1809 tatgattcctgccaaatg 1826
          ||||||||||||||||||
Db   1 tatgattcctgccaaatg 18

RESULT 38
ID   AAA60243/c standard; DNA; 18 BP.
AC   AAA60243;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 64.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; ss.
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786A-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX

DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer .
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
CC
SQ   Sequence 18 BP; 4 A; 2 C; 7 G; 5 T; 0 other;
XX

Qy   2266 gactttccaacaatgccc 2283
             ||||||||||||||||||
Db   18 GACTTTCCAACAATGCCC 1

     Query Match   0.7%;   Score 18;   DB 21;   Length 18;
     Best Local Similarity 100.0%;
     Matches 18;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

RESULT 39
AA233905
ID   AA233905 standard; DNA; 45 BP.
XX
AC   AA233905;
XX
DT   07-DEC-1999   (first entry)
XX
DE   Human PRO274 hybridisation probe.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
PD   16-SEP-1999.
XX
PF   08-MAR-1999;
XX
PR   08-MAR-1999;   99WO-US05028.

PR   10-MAR-1998;   98US-0077450.
PR   11-MAR-1998;   98US-0077632.
PR   11-MAR-1998;   98US-0077641.
PR   11-MAR-1998;   98US-0077649.
PR   12-MAR-1998;   98US-0077791.
PR   13-MAR-1998;   98US-0078004.
PR   17-MAR-1998;   98US-0040220.
PR   20-MAR-1998;   98US-0078886.
PR   20-MAR-1998;   98US-0078910.
PR   20-MAR-1998;   98US-0078936.
PR   20-MAR-1998;   98US-0078939.
PR   25-MAR-1998;   98US-0079294.
PR   26-MAR-1998;   98US-0079656.
PR   27-MAR-1998;   98US-0079664.
PR   27-MAR-1998;   98US-0079689.
PR   27-MAR-1998;   98US-0079728.
PR   30-MAR-1998;   98US-0079786.
PR   30-MAR-1998;   98US-0079920.
PR   30-MAR-1998;   98US-0079923.
PR   31-MAR-1998;   98US-0080105.
PR   31-MAR-1998;   98US-0080107.

PR   31-MAR-1998;   98US-0080165.
PR   31-MAR-1998;   98US-0080194.
PR   01-APR-1998;   98US-0080327.
PR   01-APR-1998;   98US-0080328.
PR   01-APR-1998;   98US-0080333.
PR   01-APR-1998;   98US-0080334.
PR   08-APR-1998;   98US-0081049.
PR   08-APR-1998;   98US-0081070.
PR   08-APR-1998;   98US-0081071.
PR   09-APR-1998;   98US-0081195.
PR   09-APR-1998;   98US-0081203.
PR   09-APR-1998;   98US-0081229.
PR   15-APR-1998;   98US-0081817.
PR   15-APR-1998;   98US-0081838.
PR   15-APR-1998;   98US-0081952.
PR   15-APR-1998;   98US-0081955.
PR   21-APR-1998;   98US-0082568.
PR   21-APR-1998;   98US-0082569.
PR   22-APR-1998;   98US-0082700.
PR   22-APR-1998;   98US-0082704.
PR   22-APR-1998;   98US-0082804.
PR   23-APR-1998;   98US-0082767.
PR   23-APR-1998;   98US-0082796.
PR   27-APR-1998;   98US-0083336.
PR   28-APR-1998;   98US-0083222.
PR   29-APR-1998;   98US-0083292.
PR   29-APR-1998;   98US-0083495.
PR   29-APR-1998;   98US-0083496.
PR   29-APR-1998;   98US-0083499.
PR   29-APR-1998;   98US-0083500.
PR   29-APR-1998;   98US-0083545.
PR   29-APR-1998;   98US-0083554.
PR   29-APR-1998;   98US-0083558.
PR   29-APR-1998;   98US-0083559.
PR   30-APR-1998;   98US-0083742.
PR   05-MAY-1998;   98US-0084366.
PR   06-MAY-1998;   98US-0084414.
PR   06-MAY-1998;   98US-0084441.
PR   07-MAY-1998;   98US-0084598.
PR   07-MAY-1998;   98US-0084600.
PR   07-MAY-1998;   98US-0084627.
PR   07-MAY-1998;   98US-0084637.
PR   07-MAY-1998;   98US-0084639.
PR   07-MAY-1998;   98US-0084640.
PR   07-MAY-1998;   98US-0084643.
PR   13-MAY-1998;   98US-0085323.
PR   13-MAY-1998;   98US-0085338.
PR   13-MAY-1998;   98US-0085339.
PR   15-MAY-1998;   98US-0085573.
PR   15-MAY-1998;   98US-0085579.
PR   15-MAY-1998;   98US-0085580.
PR   15-MAY-1998;   98US-0085582.
PR   15-MAY-1998;   98US-0085689.
PR   15-MAY-1998;   98US-0085697.
PR   15-MAY-1998;   98US-0085700.
PR   15-MAY-1998;   98US-0085704.
PR   18-MAY-1998;   98US-0086023.
PR   22-MAY-1998;   98US-0086392.
PR   22-MAY-1998;   98US-0086414.
PR   22-MAY-1998;   98US-0086430.
PR   22-MAY-1998;   98US-0086486.
PR   28-MAY-1998;   98US-0087098.
PR   28-MAY-1998;   98US-0087106.
PR   28-MAY-1998;   98US-0087208.
PR   30-JUL-1998;   98US-0094651.
PR   11-SEP-1998;   98US-0100038.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR   WPI; 1999-551358/46.

RESULT 40

ID   AAC78611 standard; DNA; 45 BP.
XX
AC   AAC78611;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human PRO274 hybridisation probe SEQ ID NO:17.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer; PCR primer; probe; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO2000053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.
XX
PR   08-MAR-1999; 99WO-US05028.
PR   12-MAR-1999; 99US-0123957.
PR   29-MAR-1999; 99US-0126773.
PR   21-APR-1999; 99US-0130232.
PR   28-APR-1999; 99US-0131445.
PR   14-MAY-1999; 99US-0134287.
PR   23-JUN-1999; 99US-0141037.
PR   26-JUL-1999; 99US-0145698.
PR   29-OCT-1999; 99US-0162506.
PR   30-NOV-1999; 99WO-US28313.
PR   02-DEC-1999; 99WO-US28551.
PR   02-DEC-1999; 99WO-US28565.
PR   16-DEC-1999; 99WO-US30095.
PR   30-DEC-1999; 99WO-US31243.
PR   05-JAN-2000; 99WO-US31274.
PR   06-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI   Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI   Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI   Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI   Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
DR   WPI; 2000-611443/58.
XX
PT   New secreted and transmembrane polypeptides and their polynucleotides,
PT   useful for treating blood coagulation disorders, cancers and cellular
PT   adhesion disorders -
XX
PS   Example 4; Page 184; 530pp; English.
XX
CC   The present invention describes secreted and transmembrane polypeptides
CC   and their polynucleotides. The nucleotide sequences are useful as
CC   sources of probes, primers, for chromosome mapping, and for generation
CC   of antisense sequences. They can also be used to create transgenic
CC   animals. The proteins can be used to treat a variety of diseases and
CC   disorders, depending on their function. Diseases that may be treated
CC   include blood coagulation disorders, cancers and cellular adhesion
CC   disorders. They may also be used to raise antibodies. AA233891 to
CC   AA234338, and AAY41685 to AAY41774 represent polynucleotide and
CC   polypeptide sequence given in the exemplification of the present
CC   invention.
XX
SQ   Sequence 45 BP; 8 A; 10 C; 19 G; 8 T; 0 other;

Qy   822 cattgctgctgtcaagga 839
                ||||||||||||||||||
Db   26 cattgctgctgtcaagga 43

Query Match   0.7%; Score 18; DB 20; Length 45;
Best Local Similarity   100.0%; Pred. No. 87;
Matches   18; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

PT   Novel PRO polypeptides and polynucleotides used in detection methods,
PT   to target bioactive molecules to specific cells, and to modulate
PT   cellular activities -
XX
PS   Example 4; Page 235; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC   sequence tag) sequences which encode secreted or transmembrane PRO
CC   polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC   activity. The polynucleotides and polypeptides can be used for detecting
CC   the presence of PRO polypeptides in samples, for linking bioactive
CC   molecules to cells and for modulating biological activities of cells,
CC   using the polypeptides for specific targeting. The polypeptide targeting
CC   can be used to kill the target cells, e.g. for the treatment of cancers.
CC   The polypeptide pairs provide specific targeting of bioactive molecules
CC   to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC   the isolation of the PRO polynucleotide sequences.
XX
SQ   Sequence 45 BP; 8 A; 10 C; 19 G; 8 T; 0 other;

Qy   822 cattgctgctgtcaagga 839
                ||||||||||||||||||
Db   26 cattgctgctgtcaagga 43

Query Match   0.7%; Score 18; DB 21; Length 45;
Best Local Similarity   100.0%; Pred. No. 87;
Matches   18; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 41

AAX18066/c
ID   AAX18066 standard; DNA; 177 BP.
XX
AC   AAX18066;
XX
DT   04-MAY-1999 (first entry)
XX
DE   Coding sequence for human SI binding protein SN145.
XX
KW   Gastro-intestinal transport receptor; binding protein; hSI; HP71;
KW   D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW   intestinal peptide-associated transporter; hypertension; diabetes;
KW   osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW   therapeutic agent delivery; therapy; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO9851325-A2.
XX
PD   19-NOV-1998.
XX
PF   15-MAY-1998; 98WO-US10088.
XX
PR   15-MAY-1997; 97US-0046595.
XX
PA   (CYTO-) CYTOGEN CORP.
PA   (ELAN-) ELAN CORP PLC.
XX
PI   Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
PI   Omahony DJ, Patterson CA, Singleton J;
XX
DR   WPI; 1999-009568/01.
XX
PT   New proteins that bind specifically to receptors in the
PT   gastro-intestinal tract and related nucleic acid - chimaeras and

.

PT antibodies, used to deliver therapeutic or diagnostic agents to, or
PT through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS Claim 49; Page 56; 294pp; English.
XX
CC This sequence encodes a peptide that specifically binds to the human
CC sucrose-isomaltase complex. The invention relates to purified
CC proteins (I) that bind specifically to at least one of the human
CC gastro-intestinal (GI) tract receptors human intestinal
CC peptide-associated transporter (HPT1), hPEPT1, D2H and human
CC sucrose-isomaltase complex (hsi). (I) provide active transport of
CC therapeutic agents through human and animal GI tissue (into the blood)
CC for in vivo delivery, particularly for treatment or prevention of
CC hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC migraine, or angina pectoris. Specifically they are used to deliver
CC insulin or leuprolide, but many other suitable therapeutic agents are
CC disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC antigens. (I) may also provide targeting to the GI tract. Other uses of
CC (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
SQ Sequence 177 BP; 31 A; 51 C; 53 G; 42 T; 0 other;

Qy 1766 agcagtaccacaaccagt 1783
        ||||||||||||||||||
Db 90 AGCAGTACCACAACCAGT 73

Query Match   0.7%; Score 18; DB 20; Length 177;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 42
AAT67773/c
ID AAT67773 standard; DNA; 531 BP.
XX
AC AAT67773;
XX
DT 29-JUL-1997 (first entry)
XX
DE H. Pylori cytoplasmic protein ORF 24824087.aa.
XX
KW Vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW cytoplasmic; ds.
XX
OS Helicobacter pylori.
XX
FH
FH Key   Location/Qualifiers
FT CDS   1..531
FT   /*tag= a
FT   /transl_except= (pos: 460..462, aa: Xaa)
FT   /transl_except= (pos: 520..522, aa: Xaa)
FT   /transl_except= (pos: 526..528, aa: Xaa)
FT   /note= "Xaa = Unknown"
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;
XX
DR WPI; 1997-052306/05.
DR P-PSDB; AAW20335.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 9; Page -; 1481pp; English.
XX
CC The present sequence encodes a Helicobacter pylori cytoplasmic
CC protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the
CC bacterial DNA. The sequences were analysed for ORF of at least 180
CC nucleotides, and the predicted coding regions defined by computer
CC evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences
CC of interest, particular regions can be isolated from H. pylori by
CC PCR amplification for recombinant polypeptide production, e.g. in
CC E. coli hosts.
CC Note: This DNA sequence is not reproduced in the specification and
CC has been derived from the related specification, WO9719098.
XX
SQ Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;

Qy 1710 aaagccgcttcacccttt 1727
        ||||||||||||||||||
Db 311 AAAGCCGCTTCACCCTTT 294

Query Match   0.7%; Score 18; DB 18; Length 531;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 43
AAT77453/c
ID AAT77453 standard; DNA; 531 BP.
XX
AC AAT77453;
XX
DT 11-AUG-1997 (first entry)
XX
DE H. Pylori cytoplasmic protein ORF 24824087.aa.
XX
KW Chronic gastritis; duodenal ulcer disease; activator;
KW inhibitor; bacterial life cycle; vaccine; immunisation; detection;
KW antisense; inhibition; cytoplasmic; Na+/H+ antiporter;
KW Escherichia coli; ds.
XX
OS Helicobacter pylori.
XX
FH
FH Key   Location/Qualifiers
FT CDS   1..531
FT   /*tag= a
FT   /transl_except= (pos: 460..462, aa: Xaa)
FT   /transl_except= (pos: 520..522, aa: Xaa)
FT   /transl_except= (pos: 526..528, aa: Xaa)
FT   /note= "Xaa = Unknown"
XX
PN WO9719098-A1.
XX
PD 29-MAY-1997.
XX
PF 15-NOV-1996; 96WO-US18542.

PR   17-NOV-1995;   95US-0561469.
XX
PA   (ASTR) ASTRA AB.
XX
PI   Smith DH;
XX
DR   WPI; 1997-298052/27.
DR   P-PSDB; AAW24635.
XX
PT   Helicobacter pylori nucleic acid sequences and related proteins -
     used for diagnostics and therapeutics
XX
PS   Claim 1; Page 100; 235pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori cytoplasmic
CC   protein, which was found to be homologous to Escherichia coli
CC   Na+/H+ antiporter protein following BLAST protein analysis.
CC   H. pylori has been strongly linked to chronic gastritis and
CC   duodenal ulcer disease. The nucleic acid sequences of the invention
CC   are used to evaluate compounds, especially activators or inhibitors
CC   of bacterial life cycle, for the ability to bind an H. pylori
CC   nucleic acid sequence. The nucleic acid sequences, and
CC   corresponding proteins, are also useful for generating vaccines for
CC   immunising subjects against H. pylori or for use in detecting the
CC   presence of Helicobacter species in a sample. Antisense nucleic
CC   acid sequences of these sequences are used to inhibit expression of
CC   a gene from Helicobacter species. H. pylori whole genomic DNA was
CC   isolated and nebulised to a median size of 2000 bp. Purified DNA
CC   fragments were blunt-ended and ligated to unique BstXI-linker
CC   adapters in 100-1000 fold molar excess. These linkers are
CC   complementary to the BstXI-cut pMPX vectors, while the overhang is
CC   not self-complementary. Therefore the linkers will not
CC   concatemerise nor will the cut vector re-ligate itself easily. The
CC   linker-adapter inserts were ligated to each of the 20 pMPX vectors
CC   to construct a series of shotgun subclone libraries. The purified
CC   DNA samples were then sequenced.
CC   Note: The ORF/protein reference number for this sequence was
CC   obtained from the related specification, WO9640893.
XX
SQ   Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;

Query Match   0.7%;   Score 18; DB 18; Length 531;
Best Local Similarity 100.0%;   Pred. No. 82;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1710 aaagccgcttcacccttt 1727
          ||||||||||||||||||
Db   311 AAAGCCGCTTCACCCTTT 294

RESULT 44
AA280749
ID   AA280749 standard; cDNA; 726 BP.
XX
AC   AA280749;
XX
DT   07-APR-2000 (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:833.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW   hyperplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999;   99WO-IB01062.
XX

PR   10-JUN-1998;   98US-0088801.
XX
PA   (FARB) BAYER CORP.
XX
PI   Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI   Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-087220/07.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer -
XX
PS   Claim 15; Page 462; 469pp; English.
XX
CC   AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which
CC   can be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g.
CC   carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC   can be used for developing agents for the diagnosis and treatment of
CC   disorders involving unwanted cell proliferation, such as neoplasia,
CC   dysplasia or hyperplasia.
XX
SQ   Sequence 726 BP; 149 A; 157 C; 170 G; 198 T; 52 other;

Query Match   0.7%;   Score 18; DB 21; Length 726;
Best Local Similarity 100.0%;   Pred. No. 81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   177 cctgcaggtggtggcagc 194
         ||||||||||||||||||
Db   60 cctgcaggtggtggcagc 77

RESULT 45
AAC44859
ID   AAC44859 standard; DNA; 791 BP.
XX
AC   AAC44859;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 44401.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131149.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144914.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0149368.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0149175.
PR 12-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148604.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150056.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158032.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

```
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  21-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  22-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  25-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  26-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  28-OCT-1999;  99US-0161993.
PR  29-OCT-1999;  99US-0162142.
```

```
Query Match          0.7%;   Score  18;  DB 21;   Length 791;
Best Local Similarity  100.0%;  Pred. No. 81;
  Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       2067 ggaagatggtttggaaga 2084
              ||||||||||||||||||
Db        508 ggaagatggtttggaaga 525
```

Search completed: October 28, 2001, 21:04:24
Job time: 5363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 19:32:56 : Search time 66.26 Seconds
(without alignments)
7088.453 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcgcttgctcgct..........agaaggtcagagcccagtga 2481

Scoring table: OLIGO_NUC
               Gapop 60.0 , Gapext 60.0
Searched: 324599 seqs, 94655562 residues

Word size :   8

Total number of hits satisfying chosen parameters: 376375

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19	0.8	2955	2	US-08-867-941-9	Sequence 9, Appli
2	19	0.8	2955	4	US-09-074-658-9	Sequence 9, Appli
3	19	0.8	3000	2	US-08-867-941-8	Sequence 8, Appli
4	19	0.8	3000	4	US-09-074-658-8	Sequence 8, Appli
5	19	0.8	7641	2	US-08-867-941-6	Sequence 6, Appli
6	19	0.8	7641	4	US-09-074-658-6	Sequence 6, Appli
7	18	0.8	2186	4	US-08-878-546-9	Sequence 9, Appli
8	18	0.8	2605	2	US-08-680-395-4	Sequence 4, Appli
9	17	0.7	289	2	US-08-967-101-23	Sequence 23, Appl
10	17	0.7	289	2	US-08-592-541-23	Sequence 23, Appl
11	17	0.7	289	3	US-08-124-698-23	Sequence 23, Appl
12	17	0.7	289	4	US-09-127-480-23	Sequence 23, Appl
13	17	0.7	289	4	US-08-496-841C-23	Sequence 23, Appl
14	17	0.7	304	5	US-08-611-757-20	Sequence 20, Appl
15	17	0.7	304	5	PCT-US95-05980-20	Sequence 20, Appl
16	17	0.7	1084	5	US-08-184-009-110	Sequence 110, App
17	17	0.7	1084	2	US-08-458-356-110	Sequence 110, App
18	17	0.7	1094	2	US-08-184-009-109	Sequence 109, App
19	17	0.7	1094	4	US-08-458-356-109	Sequence 109, App
20	17	0.7	1330	4	US-09-118-442-29	Sequence 29, Appl
21	17	0.7	1691	2	US-08-993-118-8	Sequence 8, Appli
22	17	0.7	1691	2	US-08-845-528C-8	Sequence 8, Appli
23	17	0.7	1816	2	US-08-951-148-2	Sequence 2, Appli
24	17	0.7	1816	3	US-09-165-234-2	Sequence 2, Appli
25	17	0.7	1816	2	US-09-274-570-2	Sequence 2, Appli
26	17	0.7	2419	1	US-07-807-043B-7	Sequence 7, Appli
27	17	0.7	2419	1	US-08-299-849B-7	Sequence 7, Appli
28	17	0.7	2419	2	US-08-142-368A-7	Sequence 7, Appli
29	17	0.7	2419	3	US-08-967-727-7	Sequence 7, Appli
30	17	0.7	2419	4	US-08-037-230D-7	Sequence 7, Appli
31	17	0.7	2420	1	US-08-465-167A-23	Sequence 23, Appl
32	17	0.7	3457	1	US-08-295-882-1	Sequence 1, Appl
33	17	0.7	3796	1	US-08-343-760A-1	Sequence 1, Appli
34	17	0.7	4488	1	US-08-441-430-1	Sequence 1, Appli
35	17	0.7	5674	1	US-07-807-043B-8	Sequence 8, Appli
36	17	0.7	5674	1	US-08-190-411A-1	Sequence 1, Appli
37	17	0.7	5674	2	US-08-299-849B-8	Sequence 8, Appli
38	17	0.7	5674	2	US-08-142-368A-8	Sequence 8, Appli
39	17	0.7	5674	3	US-08-967-727-8	Sequence 8, Appl
40	17	0.7	5674	4	US-08-037-230D-8	Sequence 8, Appli
41	17	0.7	5674	1	US-08-321-478-6	Sequence 6, Appli
42	17	0.7	6464	1	US-08-321-478-6	Sequence 6, Appli
43	17	0.7	11236	1	US-07-853-913-1	Sequence 1, Appli
44	17	0.7	29604	3	US-08-781-891-207	Sequence 207, App
45	17	0.7	68750	3	US-09-335-409-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-867-941-9
; Sequence 9, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Du, Run-Pan
; APPLICANT: Wang, Quijun
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2955 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-867-941-9

Query Match   0.8%;   Score 19;   DB 2;   Length 2955;
Best Local Similarity   100.0%;   Pred. No. 9.4;
Matches   19;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   761 tcttggtgctcaaagcaa 779
         ||||||||||||||||||

Db   591 TCTTGGTGCTCAAAGCAAA 609

RESULT 2
US-09-074-658-9
Sequence 9, Application US/09074658
Patent No. 6184371
GENERAL INFORMATION:
~ ; APPLICANT: Loosmore, Sheena M
~ ; APPLICANT: Du, Run-Pan
~ ; APPLICANT: Wang, Quijun
~ ; APPLICANT: Yang, Yan-Ping
~ ; APPLICANT: Klein, Michel H
~ ; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
~ ; NUMBER OF SEQUENCES: 78
~ ; CORRESPONDENCE ADDRESS:
~ ; ADDRESSEE: Sim & McBurney
~ ; STREET: 6th Floor, 330 University Avenue
~ ; CITY: Toronto
~ ; STATE: Ontario
~ ; COUNTRY: Canada
~ ; ZIP: M5G 1R7
~ ; COMPUTER READABLE FORM:
~ ; MEDIUM TYPE: Floppy disk
~ ; COMPUTER: IBM PC compatible
~ ; OPERATING SYSTEM: PC-DOS/MS-DOS
~ ; SOFTWARE: Patentin Release #1.0, Version #1.30
~ ; CURRENT APPLICATION DATA:
~ ; APPLICATION NUMBER: US/09/074,658
~ ; FILING DATE: 08-MAY-1998
~ ; CLASSIFICATION: 435
~ ; ATTORNEY/AGENT INFORMATION:
~ ; NAME: Stewart, Michael I
~ ; REGISTRATION NUMBER: 24,973
~ ; REFERENCE/DOCKET NUMBER: 1038-795
~ ; TELECOMMUNICATION INFORMATION:
~ ; TELEPHONE: (416) 595-1155
~ ; TELEFAX: (416) 595-1163
~ ; INFORMATION FOR SEQ ID NO: 9:
~ ; SEQUENCE CHARACTERISTICS:
~ ; LENGTH: 2955 base pairs
~ ; TYPE: nucleic acid
~ ; STRANDEDNESS: single
~ ; TOPOLOGY: linear

US-09-074-658-9

Query Match   100.0%; Score 19; DB 4; Length 2955;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 tcttggtgctcaaagcaaa 779
         |||||||||||||||||||
Db 591 TCTTGGTGCTCAAAGCAAA 609

RESULT 3
US-08-867-941-8
Sequence 8, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
~ ; APPLICANT: Loosmore, Sheena M
~ ; APPLICANT: Du, Run-Pan
~ ; APPLICANT: Wang, Quijun
~ ; APPLICANT: Yang, Yan-Ping
~ ; APPLICANT: Klein, Michel H
~ ; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
~ ; NUMBER OF SEQUENCES: 67
~ ; CORRESPONDENCE ADDRESS:
~ ; ADDRESSEE: Sim & McBurney
~ ; STREET: 6th Floor, 330 University Avenue
~ ; CITY: Toronto
~ ; STATE: Ontario
~ ; COUNTRY: Canada
~ ; ZIP: M5G 1R7
~ ; COMPUTER READABLE FORM:
~ ; MEDIUM TYPE: Floppy disk
~ ; COMPUTER: IBM PC compatible
~ ; OPERATING SYSTEM: PC-DOS/MS-DOS
~ ; SOFTWARE: Patentin Release #1.0, Version #1.30
~ ; CURRENT APPLICATION DATA:
~ ; APPLICATION NUMBER: US/08/867,941
~ ; FILING DATE: 03-JUN-1997
~ ; CLASSIFICATION: 435
~ ; ATTORNEY/AGENT INFORMATION:
~ ; NAME: Stewart, Michael I
~ ; REGISTRATION NUMBER: 24,973
~ ; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
~ ; TELECOMMUNICATION INFORMATION:
~ ; TELEPHONE: (416) 595-1155
~ ; TELEFAX: (416) 595-1163
~ ; INFORMATION FOR SEQ ID NO: 8:
~ ; SEQUENCE CHARACTERISTICS:
~ ; LENGTH: 3000 base pairs
~ ; TYPE: nucleic acid
~ ; STRANDEDNESS: single
~ ; TOPOLOGY: linear

US-08-867-941-8

Query Match   100.0%; Score 19; DB 2; Length 3000;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 761 tcttggtgctcaaagcaaa 779
         |||||||||||||||||||
Db 636 TCTTGGTGCTCAAAGCAAA 654

RESULT 4
US-09-074-658-8
Sequence 8, Application US/09074658
Patent No. 6184371
GENERAL INFORMATION:
~ ; APPLICANT: Loosmore, Sheena M
~ ; APPLICANT: Du, Run-Pan
~ ; APPLICANT: Wang, Quijun
~ ; APPLICANT: Yang, Yan-Ping
~ ; APPLICANT: Klein, Michel H
~ ; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
~ ; NUMBER OF SEQUENCES: 78
~ ; CORRESPONDENCE ADDRESS:
~ ; ADDRESSEE: Sim & McBurney
~ ; STREET: 6th Floor, 330 University Avenue
~ ; CITY: Toronto
~ ; STATE: Ontario
~ ; COUNTRY: Canada
~ ; ZIP: M5G 1R7
~ ; COMPUTER READABLE FORM:
~ ; MEDIUM TYPE: Floppy disk
~ ; COMPUTER: IBM PC compatible
~ ; OPERATING SYSTEM: PC-DOS/MS-DOS
~ ; SOFTWARE: Patentin Release #1.0, Version #1.30
~ ; CURRENT APPLICATION DATA:
~ ; APPLICATION NUMBER: US/09/074,658
~ ; FILING DATE: 08-MAY-1998
~ ; CLASSIFICATION: 435
~ ; ATTORNEY/AGENT INFORMATION:
~ ; NAME: Stewart, Michael I
~ ; REGISTRATION NUMBER: 24,973
~ ; REFERENCE/DOCKET NUMBER: 1038-795
~ ; TELECOMMUNICATION INFORMATION:
~ ; TELEPHONE: (416) 595-1155
~ ; TELEFAX: (416) 595-1163
~ ; INFORMATION FOR SEQ ID NO: 8:
~ ; SEQUENCE CHARACTERISTICS:

```
;   LENGTH: 3000 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-09-074-658-8

Qy    761 tcttggtgctcaaagcaaa 779
          |||||||||||||||||||
Db    636 TCTTGGTGCTCAAAGCAAA 654

Query Match              0.8%;  Score 19;  DB 4;  Length 3000;
Best Local Similarity  100.0%;  Pred. No. 9.4;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 5
US-08-867-941-6
; Sequence 6, Application US/0886794l
; Patent No. 5977337
; GENERAL INFORMATION:
;   APPLICANT: Loosmore, Sheena M
;   APPLICANT: Du, Run-Pan
;   APPLICANT: Wang, Quijun
;   APPLICANT: Yang, Yan-Ping
;   APPLICANT: Klein, Michel H
;   TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
;   NUMBER OF SEQUENCES: 67
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Sim & McBurney
;     STREET: 6th Floor, 330 University Avenue
;     CITY: Toronto
;     STATE: Ontario
;     COUNTRY: Canada
;     ZIP: M5G 1R7
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/867,941
;     FILING DATE: 03-JUN-1997
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Stewart, Michael I
;     REGISTRATION NUMBER: 24,973
;     REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (416) 595-1155
;     TELEFAX: (416) 595-1163
US-08-867-941-6
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear

Qy    761 tcttggtgctcaaagcaaa 779
          |||||||||||||||||||
Db   3656 TCTTGGTGCTCAAAGCAAA 3674

Query Match              0.8%;  Score 19;  DB 2;  Length 7641;
Best Local Similarity  100.0%;  Pred. No. 9.4;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 6
US-09-074-658-6
; Sequence 6, Application US/09074658
; Patent No. 6184371
```

```
;   GENERAL INFORMATION:
;     APPLICANT: Loosmore, Sheena M
;     APPLICANT: Run-Pan Du
;     APPLICANT: Quijun Wang
;     APPLICANT: Yang, Yan Ping
;     APPLICANT: Klein, Michel H
;     TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
;     NUMBER OF SEQUENCES: 78
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Sim & McBurney
;       STREET: 6th Floor, 330 University Avenue
;       CITY: Toronto
;       STATE: Ontario
;       COUNTRY: Canada
;       ZIP: M5G 1R7
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/074,658
;       FILING DATE: 08-MAY-1998
;       CLASSIFICATION: 435
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Stewart, Michael I
;       REGISTRATION NUMBER: 24,973
;       REFERENCE/DOCKET NUMBER: 1038-795
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (416) 595-1155
;       TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7641 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-09-074-658-6

Qy    761 tcttggtgctcaaagcaaa 779
          |||||||||||||||||||
Db   3656 TCTTGGTGCTCAAAGCAAA 3674

Query Match              0.8%;  Score 19;  DB 4;  Length 7641;
Best Local Similarity  100.0%;  Pred. No. 9.4;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 7
US-08-878-546-9
; Sequence 9, Application US/08878546
; Patent No. 5952463
; GENERAL INFORMATION:
;   APPLICANT: SHIBANO, YUJI
;   APPLICANT: KIKUCHI, NORIHISA
;   APPLICANT: ODA, KOHEI
;   TITLE OF INVENTION: NOVEL PROTEINASE INHIBITOR AND
;   TITLE OF INVENTION: GENE ENCODING THE INHIBITOR
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: STEINBERG, RASKIN & DAVIDSON P.C.
;     STREET: 1140 AVENUE OF THE AMERICAS
;     CITY: NEW YORK
;     STATE: NY
;     COUNTRY: USA
;     ZIP: 10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
```

Left column

```
;   APPLICATION NUMBER: US/08/878,546
;   FILING DATE: 19-JUN-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 158677/1996
;   FILING DATE: 19-JUN-1996
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 224104/1996
;   FILING DATE: 26-AUG-1996
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 48101/1997
;   FILING DATE: 03-MAR-1997
;   ATTORNEY/AGENT INFORMATION:
;   NAME: DAVIDSON, CLIFFORD M.
;   REGISTRATION NUMBER: 32,728
;   REFERENCE/DOCKET NUMBER: 382.1009
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212)-768-3800
;   TELEFAX: (212)382-2124
;   INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2186 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: STREPTOMYCES PLATENSIS
;   STRAIN: Q268
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1477..1911

US-08-878-546-9

Query Match             0.7%;  Score 18;  DB 2;  Length 2186;
Best Local Similarity 100.0%;  Pred. No. 28;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy  1602 ggtccctgggcaccctggc 1619
         |||||||||||||||||||
Db  1518 GGTCCTGGGCACCCTGGC 1535

RESULT 8
US-08-680-395-4/c
Sequence 4, Application US/08680395
Patent No. 5892010
GENERAL INFORMATION:
APPLICANT: Gray, Joe W.
APPLICANT: Collins, Colin
APPLICANT: Hwang, Soo-in
APPLICANT: Godfrey, Tony
APPLICANT: Kowbel, David
APPLICANT: Rommens, Johanna
TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
```

Right column

```
;   APPLICATION NUMBER: US/08/680,395
;   FILING DATE: 15-JUL-1996
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Bastian, Kevin L.
;   REGISTRATION NUMBER: 34,774
;   REFERENCE/DOCKET NUMBER: 023070-068900US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 576-0200
;   TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2605 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: -
;   LOCATION: 1..2605
;   OTHER INFORMATION: /note= "cDNA clone cc43 of 4 kb
;   OTHER INFORMATION: transcript"

US-08-680-395-4

Query Match             0.7%;  Score 18;  DB 2;  Length 2605;
Best Local Similarity 100.0%;  Pred. No. 28;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   398 ttccaaagtgtgtactt 415
         |||||||||||||||||
Db  1421 TTCCAAAGTGTGTACTTT 1404

RESULT 9
US-08-967-101-23/c
Sequence 23, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-967-101-23

Qy 2081 aagaggaagcagtggaa 2097
         |||||||||||||||||
Db   267 AAGAGGAAGCAGTGGAA 251

Query Match   0.7%; Score 17; DB 2; Length 289;
Best Local Similarity   100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
US-08-592-541-23/c
; Sequence 23, Application US/08592541
; Patent No. 5986054
; GENERAL INFORMATION:
;   APPLICANT: ST. GEORGE-HYSLOP, PETER H
;   APPLICANT: ROMMENS, JOHANNA M
;   APPLICANT: FRASER, PAUL E
;   TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;   TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;   NUMBER OF SEQUENCES: 183
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;   STREET: High Street Tower - 125 High Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/592,541
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   CLASSIFICATION: 800
;   NAME: Pitcher, Edmund R.
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 289 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-592-541-23

Qy 2081 aagaggaagcagtggaa 2097
         |||||||||||||||||
Db   267 AAGAGGAAGCAGTGGAA 251

Query Match   0.7%; Score 17; DB 2; Length 289;
Best Local Similarity   100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
US-09-124-698-23/c
; Sequence 23, Application US/09124698
; Patent No. 6117978
; GENERAL INFORMATION:
;   APPLICANT: ST. GEORGE-HYSLOP, PETER H
;   APPLICANT: ROMMENS, JOHANNA M

;   APPLICANT: FRASER, PAUL E
;   TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;   TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;   NUMBER OF SEQUENCES: 183
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;   STREET: High Street Tower - 125 High Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/124,698
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/592,541
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Pitcher, Edmund R.
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 289 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-09-124-698-23

Qy 2081 aagaggaagcagtggaa 2097
         |||||||||||||||||
Db   267 AAGAGGAAGCAGTGGAA 251

Query Match   0.7%; Score 17; DB 3; Length 289;
Best Local Similarity   100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
US-09-127-480-23/c
; Sequence 23, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
;   APPLICANT: ST. GEORGE-HYSLOP, PETER H
;   APPLICANT: ROMMENS, JOHANNA M
;   APPLICANT: FRASER, PAUL E
;   TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;   TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
;   NUMBER OF SEQUENCES: 183
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;   STREET: High Street Tower - 125 High Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02110
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/127,480
;   FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-127-480-23

Query Match　　　　　　0.7%; Score 17; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy　2081　aagaggaagcagtggaa　2097
　　　　　　|||||||||||||||||
Db　267　AAGAGGAAGCAGTGGAA　251

RESULT 13
US-08-496-841C-23/c
Sequence 23, Application US/0849684lC
Patent No. 6210919
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
　　ROMMENS, JOHANNA M
　　FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
　　TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby, PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,841C
FILING DATE: 28-Jun-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul F. Fehlner, Ph.D.
REGISTRATION NUMBER: 35,135
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-08-496-841C-23

Query Match

Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy　2081　aagaggaagcagtggaa　2097
　　　　　　|||||||||||||||||
Db　267　AAGAGGAAGCAGTGGAA　251

RESULT 14
US-08-611-757-20/c
Sequence 20, Application US/08611757
Patent No. 5859230
GENERAL INFORMATION:
APPLICANT: Kim, Jungsuh P.
APPLICANT: Reyes, Gregory R.
APPLICANT: Wages, John
APPLICANT: Zhang-Keck, Zhen-Yang
APPLICANT: Young, LaVonne
TITLE OF INVENTION: No. 5859230-A/No. 5859230-B/No. 5859230-C/No. 5859230-D/No.
TITLE OF INVENTION: Agents and Molecular Cloning Thereof
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,757
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,493
FILING DATE: 30-JUL-1992
APPLICATION NUMBER: US 025,396
FILING DATE: 24-FEB-1993
APPLICATION NUMBER: 08/246,985
FILING DATE: 20-MAY-1994
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: My 190 Clone D30
US-08-611-757-20

Query Match　　　　　　0.7%; Score 17; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy　82　cgccgcgagcgggccgcg　98
　　　　　|||||||||||||||||
Db　22　CGCCGCGAGCGGGCCGCG　6

RESULT 15
PCT-US95-05980-20/c
Sequence 20, Application PC/TUS9505980
GENERAL INFORMATION:
  APPLICANT:
  TITLE OF INVENTION: Non-A/Non-B/Non-C/Non-D/Non-E Hepatitis
  TITLE OF INVENTION: Agents and Molecular Cloning Thereof
  NUMBER OF SEQUENCES: 106
  CURRENT APPLICATION DATA:
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Dehlinger & Associates
    STREET: 350 Cambridge Avenue, Suite 250
    CITY: Palo Alto
    STATE: CA
    COUNTRY: USA
    ZIP: 94306
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US95/05980
    FILING DATE:
  PRIOR APPLICATION DATA:
    CLASSIFICATION:
    APPLICATION NUMBER: US 08/246,986
    FILING DATE: 20-MAY-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: Fabian, Gary R.
    REGISTRATION NUMBER: 33,875
    REFERENCE/DOCKET NUMBER: 4600-0201.49
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (415) 324-0880
    TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 20:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 304 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
  MOLECULE TYPE: DNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
    INDIVIDUAL ISOLATE: My 190 Clone D30

Query Match
Best Local Similarity   100.0%; Score 17; DB 5; Length 304;
Matches   17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 cgccgcgaggcgccgcg 98
        |||||||||||||||||
Db 22 CGCCGCGAGGCGGCGCG 6

RESULT 16
US-08-184-009-110
Sequence 110, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
  APPLICANT: Paoletti, Enzo
  APPLICANT: Tartaglia, James
  APPLICANT: Cox, William I.
  TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
  NUMBER OF SEQUENCES: 217
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Curtis, Morris & Safford
    STREET: 530 Fifth Avenue

    CITY: New York
    STATE: NY
    COUNTRY: USA
    ZIP: 10036
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/184,009
    FILING DATE: 19-JAN-1994
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Frommer, William S.
    REGISTRATION NUMBER: 25,506
    REFERENCE/DOCKET NUMBER: 454310-2530
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212) 840-3333
    TELEFAX: (212) 840-0712
    TELEX: 4250660CURTMS
INFORMATION FOR SEQ ID NO: 110:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1084 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
US-08-184-009-110

Query Match
Best Local Similarity   100.0%; Score 17; DB 2; Length 1084;
Matches   17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1602 ggtccttgggcaccctgg 1618
          ||||||||||||||||||
Db   292 GGTCCTTGGGCACCCTGG 308

RESULT 17
US-08-458-356-110
Sequence 110, Application US/08458356
Patent No. 5942235
GENERAL INFORMATION:
  APPLICANT: Paoletti, Enzo
  APPLICANT: Tartaglia, James
  APPLICANT: Cox, William I.
  TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
  NUMBER OF SEQUENCES: 217
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Curtis, Morris & Safford
    STREET: 530 Fifth Avenue
    CITY: New York
    STATE: NY
    COUNTRY: USA
    ZIP: 10036
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/458,356
    FILING DATE: 02-JUN-1995
  PRIOR APPLICATION DATA:
    CLASSIFICATION: 424
    APPLICATION NUMBER: US 08/184,009
    FILING DATE: 19-JAN-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: Frommer, William S.
    REGISTRATION NUMBER: 25,506
    REFERENCE/DOCKET NUMBER: 454310-2530

```
                                                      TELECOMMUNICATION INFORMATION:
                                                         TELEPHONE: (212) 840-3333
                                                         TELEFAX: (212) 840-0712
                                                         TELEX: 425066CURTMS
                                                   INFORMATION FOR SEQ ID NO: 110:
                                                   SEQUENCE CHARACTERISTICS:
                                                         LENGTH: 1084 base pairs
                                                         TYPE: nucleic acid
                                                         STRANDEDNESS: single
                                                         TOPOLOGY: linear
                                                         MOLECULE TYPE: cDNA
                                                   US-08-458-356-110

Qy  1602 ggtcctgggcaccctgg 1618
         |||||||||||||||||
Db   292 GGTCCTGGGCACCCTGG 308
```

```
Query Match            100.0%;  Score 17;  DB 2;  Length 1084;
Best Local Similarity  100.0%;  Pred. No. 87;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
RESULT 18
US-08-184-009-109
Sequence 109, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
   APPLICANT: Paoletti, Enzo
   APPLICANT: Tartaglia, James
   APPLICANT: Cox, William I.
   TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
   NUMBER OF SEQUENCES: 217
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: Curtis, Morris & Safford
   STREET: 530 Fifth Avenue
   CITY: New York
   STATE: NY
   COUNTRY: USA
   ZIP: 10036
   COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/184,009
   FILING DATE: 19-JAN-1994
   CLASSIFICATION: 435
   ATTORNEY/AGENT INFORMATION:
   NAME: Frommer, William S.
   REGISTRATION NUMBER: 25,506
   REFERENCE/DOCKET NUMBER: 454310-2530
   TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 840-3333
   TELEFAX: (212) 840-0712
   TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
   LENGTH: 1094 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: single
   TOPOLOGY: linear
   MOLECULE TYPE: cDNA
US-08-184-009-109
```

```
Qy  1602 ggtcctgggcaccctgg 1618
         |||||||||||||||||
Db   314 GGTCCTGGGCACCCTGG 330
```

```
Query Match            100.0%;  Score 17;  DB 2;  Length 1094;
Best Local Similarity  100.0%;  Pred. No. 87;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
RESULT 19
US-08-458-356-109
Sequence 109, Application US/08458356
Patent No. 594235
GENERAL INFORMATION:
   APPLICANT: Paoletti, Enzo
   APPLICANT: Tartaglia, James
   APPLICANT: Cox, William I.
   TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
   NUMBER OF SEQUENCES: 217
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: Curtis, Morris & Safford
   STREET: 530 Fifth Avenue
   CITY: New York
   STATE: NY
   COUNTRY: USA
   ZIP: 10036
   COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/458,356
   FILING DATE: 02-JUN-1995
   CLASSIFICATION: 424
   PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 08/184,009
   FILING DATE: 19-JAN-1994
   ATTORNEY/AGENT INFORMATION:
   NAME: Frommer, William S.
   REGISTRATION NUMBER: 25,506
   REFERENCE/DOCKET NUMBER: 454310-2530
   TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 840-3333
   TELEFAX: (212) 840-0712
   TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
   LENGTH: 1094 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: single
   TOPOLOGY: linear
   MOLECULE TYPE: cDNA
US-08-458-356-109
```

```
Qy  1602 ggtcctgggcaccctgg 1618
         |||||||||||||||||
Db   314 GGTCCTGGGCACCCTGG 330
```

```
Query Match            100.0%;  Score 17;  DB 2;  Length 1094;
Best Local Similarity  100.0%;  Pred. No. 87;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
RESULT 20
US-09-118-442-29/c
Sequence 29, Application US/09118442B
Patent No. 6197561
GENERAL INFORMATION:
   APPLICANT: Martino-Catt, Susan J.
   APPLICANT: Wang, Hongyu
   APPLICANT: Beach, Larry R.
   APPLICANT: Wang, Xun
   APPLICANT: Bowen, Benjamin A.
   TITLE OF INVENTION: Genes Controlling Phytate Metabolism in
   TITLE OF INVENTION: Plants and Uses Thereof
   FILE REFERENCE: 0706
   CURRENT APPLICATION NUMBER: US/09/118,442B
```

```
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Zea mays
US-09-118-442-29

Qy    1473 catccgatgaagattc 1489
           ||||||||||||||||
Db     493 CATCCGATGAAGATTC 477

Query Match   Best Local Similarity 100.0%;   Score 17;  DB 4;  Length 1330;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
US-09-118-442-29

RESULT 21
US-08-993-118-8
Sequence 8, Application US/08993118
Patent No. 5597872
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,118
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; APPLICATION NUMBER: US/08/845,528
; FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1691 base pairs
; TYPE: nucleotides
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
US-08-993-118-8

Query Match   0.7%;   Score 17;  DB 2;  Length 1691;
```

```
Best Local Similarity 100.0%;   Pred. No. 87;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1602 ggtcctgggcaccctgg 1618
           |||||||||||||||||
Db     317 GGTCCTGGGCACCCTGG 333

RESULT 22
US-08-845-528C-8
Sequence 8, Application US/08845528C
Patent No. 6027924
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,528C
; FILING DATE: April 25, 1997
; CLASSIFICATION: 4335
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary Anne Schofield
; REGISTRATION NUMBER: 36,669
; REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1691 base pairs
; TYPE: nucleotides
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
US-08-845-528C-8

Query Match   Best Local Similarity 100.0%;   Score 17;  DB 3;  Length 1691;
Matches 17;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1602 ggtcctgggcaccctgg 1618
           |||||||||||||||||
Db     317 GGTCCTGGGCACCCTGG 333

RESULT 23
US-08-951-148-2
Sequence 2, Application US/08951148
Patent No. 5871973
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: CELL DIVISION REGULATORS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
```

```
              ADDRESSEE: Incyte Pharmaceuticals, Inc.
              STREET: 3174 Porter Drive
              CITY: Palo Alto
              STATE: CA
              COUNTRY: USA
              ZIP: 94304
       COMPUTER READABLE FORM:
              MEDIUM TYPE: Diskette
              COMPUTER: IBM Compatible
              OPERATING SYSTEM: DOS
              SOFTWARE: FastSEQ for Windows Version 2.0
       CURRENT APPLICATION DATA:
              APPLICATION NUMBER: US/08/951,148
              FILING DATE:
              CLASSIFICATION: 514
       PRIOR APPLICATION DATA:
              APPLICATION NUMBER:
              FILING DATE:
       ATTORNEY/AGENT INFORMATION:
              NAME: Billings, Lucy J.
              REGISTRATION NUMBER: 36,749
              REFERENCE/DOCKET NUMBER: PF-0407 US
       TELECOMMUNICATION INFORMATION:
              TELEPHONE: 650-855-0555
              TELEFAX: 650-845-4166
              TELEX:
INFORMATION FOR SEQ ID NO: 2:
       SEQUENCE CHARACTERISTICS:
              LENGTH: 1816 base pairs
              TYPE: nucleic acid
              STRANDEDNESS: single
              TOPOLOGY: linear
       IMMEDIATE SOURCE:
              LIBRARY: SPLNFZT01
              CLONE: 26459
US-08-951-148-2

Qy   754 ggaaacttcttggtgct 770
         |||||||||||||||||
Db   717 GGAAACTTCTTGGTGCT 733

Query Match
Best Local Similarity   100.0%;   Score 17;   DB 2;   Length 1816;
Matches  17;  Conservative  0;  Pred. No. 87;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 24
US-09-165-234-2
; Sequence 2, Application US/09165234
; Patent No. 5928899
; GENERAL INFORMATION:
       APPLICANT: Hillman, Jennifer L.
       APPLICANT: Bandman, Olga
       APPLICANT: Lal, Preeti
       APPLICANT: Shah, Purvi
       TITLE OF INVENTION: CELL DIVISION REGULATORS
       NUMBER OF SEQUENCES: 10
       CORRESPONDENCE ADDRESS:
              ADDRESSEE: Incyte Pharmaceuticals, Inc.
              STREET: 3174 Porter Drive
              CITY: Palo Alto
              STATE: CA
              COUNTRY: USA
              ZIP: 94304
       COMPUTER READABLE FORM:
              MEDIUM TYPE: Diskette
              COMPUTER: IBM Compatible
              OPERATING SYSTEM: DOS
              SOFTWARE: FastSEQ for Windows Version 2.0
       CURRENT APPLICATION DATA:
              APPLICATION NUMBER: US/09/165,234
              FILING DATE:
```

```
              CLASSIFICATION:
       PRIOR APPLICATION DATA:
              APPLICATION NUMBER: 08/951,148
              FILING DATE:
       ATTORNEY/AGENT INFORMATION:
              NAME: Billings, Lucy J.
              REGISTRATION NUMBER: 36,749
              REFERENCE/DOCKET NUMBER: PF-0407 US
       TELECOMMUNICATION INFORMATION:
              TELEPHONE: 650-855-0555
              TELEFAX: 650-845-4166
              TELEX:
INFORMATION FOR SEQ ID NO: 2:
       SEQUENCE CHARACTERISTICS:
              LENGTH: 1816 base pairs
              TYPE: nucleic acid
              STRANDEDNESS: single
              TOPOLOGY: linear
       IMMEDIATE SOURCE:
              LIBRARY: SPLNFZT01
              CLONE: 26459
US-09-165-234-2

Qy   754 ggaaacttcttggtgct 770
         |||||||||||||||||
Db   717 GGAAACTTCTTGGTGCT 733

Query Match
Best Local Similarity   100.0%;   Score 17;   DB 2;   Length 1816;
Matches  17;  Conservative  0;  Pred. No. 87;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 25
US-09-274-570-2
; Sequence 2, Application US/09274570
; Patent No. 6121019
; GENERAL INFORMATION:
       APPLICANT: Hillman, Jennifer L.
       APPLICANT: Bandman, Olga
       APPLICANT: Lal, Preeti
       APPLICANT: Shah, Purvi
       TITLE OF INVENTION: CELL DIVISION REGULATORS
       NUMBER OF SEQUENCES: 10
       CORRESPONDENCE ADDRESS:
              ADDRESSEE: Incyte Pharmaceuticals, Inc.
              STREET: 3174 Porter Drive
              CITY: Palo Alto
              STATE: CA
              COUNTRY: USA
              ZIP: 94304
       COMPUTER READABLE FORM:
              MEDIUM TYPE: Diskette
              COMPUTER: IBM Compatible
              OPERATING SYSTEM: DOS
              SOFTWARE: FastSEQ for Windows Version 2.0
       CURRENT APPLICATION DATA:
              APPLICATION NUMBER: US/09/274,570
              FILING DATE:
              CLASSIFICATION:
       PRIOR APPLICATION DATA:
              APPLICATION NUMBER: 08/951,148
              FILING DATE:
       ATTORNEY/AGENT INFORMATION:
              NAME: Billings, Lucy J.
              REGISTRATION NUMBER: 36,749
              REFERENCE/DOCKET NUMBER: PF-0407 US
       TELECOMMUNICATION INFORMATION:
              TELEPHONE: 650-855-0555
              TELEFAX: 650-845-4166
              TELEX:
INFORMATION FOR SEQ ID NO: 2:
       SEQUENCE CHARACTERISTICS:
```

;
; LENGTH: 1816 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: SPLNFZT01
;   CLONE: 26459
;
US-09-274-570-2

Qy   754 ggaaacttcttggtgct 770
          |||||||||||||||||
Db   717 GGAAACTCTTGGTGCT 733

Query Match   0.7%;   Score 17; DB 3; Length 1816;
Best Local Similarity 100.0%;   Pred. No. 87;
Matches   17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 26
US-07-807-043B-7
; Sequence 7, Application US/0780704 3B
; Patent No. 5342774
; GENERAL INFORMATION:
~   APPLICANT: Boon, Thierry, Van den Eynde, Benot
~   TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
~   TITLE OF INVENTION: Rejection Antigens and Uses Thereof
~   NUMBER OF SEQUENCES: 16
~   CORRESPONDENCE ADDRESS:
~   ADDRESSEE: Felfe & Lynch
~   STREET: 805 Third Avenue
~   CITY: New York City
~   STATE: New York
~   ZIP: 10022
~   COMPUTER READABLE FORM:
~   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
~   COMPUTER: IBM
~   OPERATING SYSTEM: PC-DOS
~   SOFTWARE: Wordperfect
~   CURRENT APPLICATION DATA:
~   APPLICATION NUMBER: US/07/807,043B
~   FILING DATE: 19911212
~   CLASSIFICATION: 424
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/764,364
~   FILING DATE: 23-SEPTEMBER-1991
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/728,838
~   FILING DATE: 9-JULY-1991
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/705,702
~   FILING DATE: 23-May-1991
~   ATTORNEY/AGENT INFORMATION:
~   NAME: Hanson, No. 5342774man D.
~   REGISTRATION NUMBER: 30,946
~   REFERENCE/DOCKET NUMBER: LUD 253.3
~   TELECOMMUNICATION INFORMATION:
~   TELEPHONE: (212) 688-9200
~   TELEFAX: (212) 838-3884
~ INFORMATION FOR SEQ ID NO: 7:
~   SEQUENCE CHARACTERISTICS:
~   LENGTH: 2419 base pairs
~   TYPE: NUCLEIC ACID
~   STRANDEDNESS: singular
~   TOPOLOGY: linear
~   MOLECULE TYPE: genomic DNA
US-07-807-043B-7

Query Match   0.7%;   Score 17; DB 1; Length 2419;
Best Local Similarity 100.0%;   Pred. No. 87;
Matches   17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1602 ggtcctgggcaccctgg 1618
          |||||||||||||||||
Db   739 GGTCCTGGGCACCCTGG 755

RESULT 27
US-08-299-849B-7
; Sequence 7, Application US/0829984 9B
; Patent No. 5612201
; GENERAL INFORMATION:
~   APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
~   APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
~   APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
~   APPLICANT: Chomez, Patrick
~   TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
~   TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
~   NUMBER OF SEQUENCES: 48
~   CORRESPONDENCE ADDRESS:
~   ADDRESSEE: Felfe & Lynch
~   STREET: 805 Third Avenue
~   CITY: New York City
~   STATE: New York
~   ZIP: 10022
~   COMPUTER READABLE FORM:
~   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
~   COMPUTER: IBM
~   OPERATING SYSTEM: PC-DOS
~   SOFTWARE: Wordperfect
~   CURRENT APPLICATION DATA:
~   APPLICATION NUMBER: US/08/299,849B
~   FILING DATE: 1-SEPTEMBER-1994
~   CLASSIFICATION: 435
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 08/037,230
~   FILING DATE: 26-MARCH-1993
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: PCT/US92/04354
~   FILING DATE: 22-MAY-1992
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/807,043
~   FILING DATE: 12-DECEMBER-1991
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/764,364
~   FILING DATE: 23-SEPTEMBER-1991
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/728,838
~   FILING DATE: 9-JULY-1991
~   PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: 07/705,702
~   FILING DATE: 23-May-1991
~   ATTORNEY/AGENT INFORMATION:
~   NAME: Hanson, No. 5612201man D.
~   REGISTRATION NUMBER: 30,946
~   REFERENCE/DOCKET NUMBER: LUD 5355
~   TELECOMMUNICATION INFORMATION:
~   TELEPHONE: (212) 688-9200
~   TELEFAX: (212) 838-3884
~ INFORMATION FOR SEQ ID NO: 7:
~   SEQUENCE CHARACTERISTICS:
~   LENGTH: 2419 base pairs
~   TYPE: nucleic acid
~   STRANDEDNESS: single
~   TOPOLOGY: linear
~   MOLECULE TYPE: genomic DNA
US-08-299-849B-7

Query Match   0.7%;   Score 17; DB 1; Length 2419;
Best Local Similarity 100.0%;   Pred. No. 87;
Matches   17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1602 ggtcctgggcaccctgg 1618
          |||||||||||||||||
Db   739 GGTCCTGGGCACCCTGG 755

RESULT 28
US-08-142-368A-7
; Sequence 7, Application US/0814236BA
; Patent No. 5925729
; GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   ZIP: 10022
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;   COMPUTER: IBM
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/142,368A
;   FILING DATE: 02-MAY-1994
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US92/04354
;   FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/807,043
;   FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/764,364
;   FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/728,838
;   FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/705,702
;   FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hanson, No. 592572man D.
;   REGISTRATION NUMBER: 30,946
;   REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 688-9200
;   TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2419 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: genomic DNA

Query Match   0.7%; Score 17; DB 2; Length 2419;
Best Local Similarity   100.0%; Pred. No. 87;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1602 ggtcctgggcaccctgg 1618
            |||||||||||||||||
Db   739 GGTCCTGGGCACCCTGG 755

RESULT 29
US-08-967-727-7
; Sequence 7, Application US/08967727
; Patent No. 6025474

; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   ZIP: 10022
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;   COMPUTER: IBM
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/967,727
;   FILING DATE:
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/037,230
;   FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US92/04354
;   FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/807,043
;   FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/764,365
;   FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/728,838
;   FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/705,702
;   FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: Hanson, No. 602547man D.
;   REGISTRATION NUMBER: 30,946
;   REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 688-9200
;   TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2419 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: genomic DNA
US-08-967-727-7

Query Match   0.7%; Score 17; DB 3; Length 2419;
Best Local Similarity   100.0%; Pred. No. 87;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1602 ggtcctgggcaccctgg 1618
            |||||||||||||||||
Db   739 GGTCCTGGGCACCCTGG 755

RESULT 30
US-08-037-230D-7
; Sequence 7, Application US/08037230D
; Patent No. 6235525
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof

; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,230D
; FILING DATE: 26-MARCH-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, NO. 623552man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-037-230D-7

Query Match　　0.7%;　　Score 17;　DB 4;　Length 2419;
Best Local Similarity 100.0%;　Pred. No. 87;
Matches 17;　Conservative 0;　Mismatches 0;　Indels 0;　Gaps 0;

Qy 1602 ggtcctgggcaccctgg 1618
         |||||||||||||||||
Db 739 GGTCCTGGGCACCCTGG 755

RESULT 31
US-08-465-167A-23
; Sequence 23, Application US/0846516 7A
; Patent No. 5750395
; GENERAL INFORMATION:
; APPLICANT: Fikes, John D.
; APPLICANT: Livingston, Brian D.
; APPLICANT: Sette, Alessandro D.
; APPLICANT: Sidney, John C.
; TITLE OF INVENTION: DNA ENCODING MAGE-1 C-TERMINAL
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES (as amended)
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco

; STATE: CA
; COUNTRY: USA
; ZIP: 98111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,167A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/103,623
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 14137-60-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2420 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 626..1552
US-08-465-167A-23

Query Match　　0.7%;　　Score 17;　DB 1;　Length 2420;
Best Local Similarity 100.0%;　Pred. No. 87;
Matches 17;　Conservative 0;　Mismatches 0;　Indels 0;　Gaps 0;

Qy 1602 ggtcctgggcaccctgg 1618
         |||||||||||||||||
Db 739 GGTCCTGGGCACCCTGG 755

RESULT 32
US-08-295-882-1/c
; Sequence 1, Application US/08295882
; Patent No. 5569833
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHOD FOR ENHANCING
; TITLE OF INVENTION: PLANT PRECOCITY AND/OR REDUCING THE
; TITLE OF INVENTION: STORED NITRATE CONTENT OF A PLANT
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie, Parker & Hale
; STREET: P.O. Box 7068
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91109-7068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORD PERFECT 5.1 release
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,882
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00222
; FILING DATE: March 5, 1993
; APPLICATION NUMBER: 92 02658

RESULT 33
US-08-343-760A-1
Sequence 1, Application US/08343760A
Patent No. 5679783

GENERAL INFORMATION:
  APPLICANT: De Robertis, Edward M
  APPLICANT: Sasai, Yoshiki
  TITLE OF INVENTION: Tissue Differentiation Affecting
  TITLE OF INVENTION: Factor and Composition
  NUMBER OF SEQUENCES: 3
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Majestic, Parsons, Siebert & Hsue
    STREET: Four Embarcadero Center, Suite 1450
    CITY: San Francisco
    STATE: CA
    COUNTRY: USA
    ZIP: 94111
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/343,760A
    FILING DATE: 22-NOV-1994
    CLASSIFICATION: 536
  ATTORNEY/AGENT INFORMATION:
    NAME: Siebert, J. Suzanne
    REGISTRATION NUMBER: 28,758
    REFERENCE/DOCKET NUMBER: 3100.1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (415) 363-5556
    TELEFAX: (415) 362-5418
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:

Query Match 0.7%; Score 17; DB 1; Length 3457;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   859 gaaggaagaagagatttt 875
         ||||||||||||||||||
Db   94 GAAGGAAGAAGAGATTTT 78

FILING DATE: March 5, 1992
ATTORNEY/AGENT INFORMATION:
  NAME: Prout, D. Bruce
  REGISTRATION NUMBER: 20,958
  REFERENCE/DOCKET NUMBER: 27209/DBP
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 3457 base pairs
    TYPE: nucleotide with corresponding
    TYPE: protein
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: cDNA
  ORIGINAL SOURCE:
  ORIGINAL SOURCE: (1.vi.A) ORGANISM: Nicotiana tabacum
  ORIGINAL SOURCE: (1.vi.B) CELL LINE: N. tabacum cv. Xanthi
  ORIGINAL SOURCE: XHFD 8
  IMMEDIATE SOURCE: leaf
  FEATURE:
    NAME/KEY: Nitrate reductase
    LOCATION: from 1 to 143 bp: Leader
    LOCATION: non translated 5 sequence (leader)
    LOCATION: from 144 to 2855 bp: coding sequence
    LOCATION: for nitrate reductase apoenzyme
    LOCATION: from 2856 to 3457 bp: non translated
    LOCATION: 3 sequence
US-08-295-882-1

RESULT 34
US-08-441-430-1
Sequence 1, Application US/08441430
Patent No. 5681942

GENERAL INFORMATION:
  APPLICANT: Buchwald, Manuel
  APPLICANT: Strathdee, Craig A.
  APPLICANT: Wevrick, Rachel
  APPLICANT: Mathew, Christopher George Porter
  TITLE OF INVENTION: Fanconi Anemia Type C Gene
  NUMBER OF SEQUENCES: 73
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Richard J. Polley, Esq.
    ADDRESSEE: Klarquist, Sparkman, Campbell, Leigh &
    ADDRESSEE: Whinston, LLP
    STREET: 121 S.W. Salmon, Suite 1600
    CITY: Portland
    STATE: Oregon
    COUNTRY: U.S.A.
    ZIP: 97204
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Disk, 3+-inch
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: MS DOS
    SOFTWARE: WordPerfect 5.1/ASCII Text File
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/441,430
    FILING DATE: May 15, 1995
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: U.S. 07/876,285
    FILING DATE: April 29, 1992
    APPLICATION NUMBER: U.S. 07/918,313
    FILING DATE: July 21, 1992
    APPLICATION NUMBER: U.S. 08/003,963
    FILING DATE: January 15, 1993
  ATTORNEY/AGENT INFORMATION:
    NAME: Richard J. Polley, Esq.
    REGISTRATION NUMBER: 28,107
    REFERENCE/DOCKET NUMBER: 3812-42824
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (503) 226-7391
    TELEFAX: (503) 228-9446
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 4488 base pairs
      TYPE: Nucleic Acid
      STRANDEDNESS: Double stranded
      TOPOLOGY: Linear
      MOLECULE TYPE: cDNA to mRNA
      ANTI-SENSE: NO
      HYPOTHETICAL: NO
    ORIGINAL SOURCE:
      ORGANISM: Homo sapiens
    IMMEDIATE SOURCE:
      LIBRARY: Human cDNA

LENGTH: 3796 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-343-760A-1

Query Match 100.0%; Score 17; DB 1; Length 3796;
Best Local Similarity 0.7%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   220 ctctacgtctctccga 236
         ||||||||||||||||
Db 1036 CTCTACGTCTTCTCCGA 1052

```
; POSITION IN GENOME:      (of corresponding genomic gene)
;   CHROMOSOME/SEGMENT:    9q
;   MAP POSITION: 22.3
;   UNITS:
US-08-441-430-1

Query Match             0.7%;  Score 17;  DB 1;  Length 4488;
Best Local Similarity   100.0%;  Pred. No. 87;
Matches   17;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   2055 tgaagccaccctggaag 2071
          |||||||||||||||||
Db   4189 TGAAGCCACCCTGGAAG 4205

RESULT 35
US-07-807-043B-8
; Sequence 8, Application US/0780704 3B
; Patent No. 5342774
;   GENERAL INFORMATION:
;     APPLICANT: Boon, Thierry, Van den Eynde, Beno t
;     TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
;     TITLE OF INVENTION: Rejection Antigens and Uses Thereof
;     NUMBER OF SEQUENCES: 16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Felfe & Lynch
;       STREET: 805 Third Avenue
;       CITY: New York City
;       STATE: New York
;       ZIP: 10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;       COMPUTER: IBM
;       OPERATING SYSTEM: PC-DOS
;       SOFTWARE: Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/07/807,043B
;       FILING DATE: 19911212
;       CLASSIFICATION: 424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/764,364
;       FILING DATE: 23-SEPTEMBER-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/728,838
;       FILING DATE: 9-JULY-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/705,702
;       FILING DATE: 23-May-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Hanson, No. 5342774man D.
;       REGISTRATION NUMBER: 30,946
;       REFERENCE/DOCKET NUMBER: LUD 253.3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (212) 688-9200
;       TELEFAX: (212) 838-3884
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 5674 base pairs
;       TYPE: NUCLEIC ACID
;       STRANDEDNESS: singular
;       TOPOLOGY: linear
;     MOLECULE TYPE: genomic DNA
;     FEATURE:
;       NAME/KEY: MAGE-1 gene
US-07-807-043B-8

Query Match             0.7%;  Score 17;  DB 1;  Length 5674;
Best Local Similarity   100.0%;  Pred. No. 87;
Matches   17;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1602 ggtcctggcaccctgg 1618
          |||||||||||||||||
Db   3994 GGTCCTGGGCACCCTGG 4010

RESULT 36
US-08-190-411A-1
; Sequence 1, Application US/08190411A
; Patent No. 5541104
;   GENERAL INFORMATION:
;     APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
;     APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
;     APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry;
;     APPLICANT: Old, Lloyd J.
;     TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
;     TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-
;     TITLE OF INVENTION: AND MAGE-1 DERIVED IMMUNOGENIC PEPTIDES
;     NUMBER OF SEQUENCES: 4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Felfe & Lynch
;       STREET: 805 Third Avenue
;       CITY: New York City
;       STATE: New York
;       ZIP: 10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;       COMPUTER: IBM
;       OPERATING SYSTEM: PC-DOS
;       SOFTWARE: Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/190,411A
;       FILING DATE: 01-FEBRUARY-1994
;       CLASSIFICATION: 436
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 037,230
;       FILING DATE: 26-MARCH-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: PCT/US92/04354
;       FILING DATE: 22-MAY-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/807,043
;       FILING DATE: 12-DECEMBER-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/764,364
;       FILING DATE: 23-SEPTEMBER-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/728,838
;       FILING DATE: 9-JULY-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/705,702
;       FILING DATE: 23-MAY-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Hanson, No. 554104man D.
;       REGISTRATION NUMBER: 30,946
;       REFERENCE/DOCKET NUMBER: LUD 5354
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (212) 688-9200
;       TELEFAX: (212) 838-3884
;   INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 5674 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;     MOLECULE TYPE: genomic DNA
;     FEATURE:
;       NAME/KEY: MAGE-1 gene
US-08-190-411A-1

Query Match             0.7%;  Score 17;  DB 1;  Length 5674;
Best Local Similarity   100.0%;  Pred. No. 87;
Matches   17;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

Left Column

```
Qy      1602 ggtcctgggcaccctgg 1618
             |||||||||||||||||
Db      3994 GGTCCTGGGCACCCTGG 4010
```

RESULT 37
US-08-299-849B-8
Sequence 8, Application US/0829984B9
Patent No. 5612201

; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612012man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
US-08-299-849B-8
```

Query Match
Best Local Similarity 100.0%; Pred. No. 87; DB 1; Length 5674;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1602 ggtcctgggcaccctgg 1618
```

Right Column

```
Db      3994 GGTCCTGGGCACCCTGG 4010
             |||||||||||||||||
```

RESULT 38
US-08-560-024-1
Sequence 1, Application US/08560024
Patent No. 5843448

; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
; APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry;
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-
; TITLE OF INVENTION: AND MAGE-1 DERIVED IMMUNOGENIC PEPTIDES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,024
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,411
; FILING DATE: 01-FEBRUARY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5843448man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5354
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene
US-08-560-024-1
```

Query Match
Best Local Similarity 100.0%; Pred. No. 87; DB 2; Length 5674;

Score 17; Score 17;

Qy 1602 ggtcctgggcaccctgg 1618
|||||||||||||||||||
Db 3994 GGTCCTGGGCACCCTGG 4010

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 39
US-08-142-368A-8
Sequence 8, Application US/08142368A
Patent No. 5925729
GENERAL INFORMATION:
APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
TITLE OF INVENTION: Rejection Antigens and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/142,368A
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE: 23-May-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 592572929man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5253.4-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 5674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-1 gene

Query Match 100.0%; Score 17; DB 2; Length 5674;
Best Local Similarity 0.7%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
US-08-142-368A-8

Qy 1602 ggtcctgggcaccctgg 1618
|||||||||||||||||||
Db 3994 GGTCCTGGGCACCCTGG 4010

---

Db 3994 GGTCCTGGGCACCCTGG 4010

RESULT 40
US-08-967-727-8
Sequence 8, Application US/08967727
Patent No. 6025474
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,727
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,365
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 602574man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 5674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-1 gene

Query Match 100.0%; Score 17; DB 3; Length 5674;
Best Local Similarity 0.7%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
US-08-967-727-8

Qy 1602 ggtcctgggcaccctgg 1618
|||||||||||||||||||
Db 3994 GGTCCTGGGCACCCTGG 4010

RESULT 41
US-08-037-230D-8
Sequence 8, Application US/08037230D
Patent No. 6235525
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/037,230D
FILING DATE: 26-MARCH-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 623552 5man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 5674 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: MAGE-1 gene

US-08-037-230D-8

Query Match
Best Local Similarity 100.0%;   Score 17;   DB 4;   Length 5674;
Matches 17;   Conservative 0;   Pred. No. 87;
Mismatches 0;   Indels 0;   Gaps 0;

Qy 1602 ggtcctgggcaccctgg 1618
        |||||||||||||||||
Db 3994 GGTCCTGGGCACCCTGG 4010

RESULT 42
US-08-321-478-6
Sequence 6, Application US/08321478
Patent No. 5527677
GENERAL INFORMATION:
APPLICANT: DEGUCHI, Takeo
APPLICANT: KINOSHITA, Moritoshi
APPLICANT: KATSURAGI, Kiyonori
APPLICANT: SHIN, Sadahito
TITLE OF INVENTION: HUMAN ARYLAMINE N-ACETYLTRANSFERASE
TITLE OF INVENTION: GENES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/321,478
FILING DATE: 11-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/038,667
FILING DATE: 23-MAR-1993
APPLICATION NUMBER: JP 64669/1992
FILING DATE: 23-MAR-1992
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6464 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 723..1595
FEATURE:
NAME/KEY: exon
LOCATION: 717..1936
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1794..1799
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1800..1805
US-08-321-478-6

Query Match
Best Local Similarity 100.0%;   Score 17;   DB 1;   Length 6464;
Matches 17;   Conservative 0;   Pred. No. 87;
Mismatches 0;   Indels 0;   Gaps 0;

Qy 982 aggtaccaaggaaaggc 998
       |||||||||||||||||
Db 5916 AGGTACCAAGGAAAGGC 5932

RESULT 43
US-07-853-913-1
Sequence 1, Application US/0785913913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
APPLICANT: Lendahl, Urban
TITLE OF INVENTION: Nestin Expression As An Indicator of
TITLE OF INVENTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11236 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-07-853-913-1

Qy    1827 ccttcaggaaggggctg 1843
           |||||||||||||||||
Db    3466 CCTTCAGGAAGGGGCTG 3482

Query Match       0.7%;  Score 17;  DB 1;  Length 11236;
Best Local Similarity 100.0%;  Pred. No. 87;
Matches   17;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 44
US-08-781-891-207
Sequence 207, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Milligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
TITLE OF INVENTION: WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 29604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-781-891-207

Qy     475 aaaggaatagaactggc 491
           |||||||||||||||||
Db   14826 AAAGGAATAGAACTGGC 14842

Query Match       0.7%;  Score 17;  DB 3;  Length 29604;
Best Local Similarity 100.0%;  Pred. No. 86;
Matches   17;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

RESULT 45
US-09-335-409-1/c
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum

US-09-335-409-1

Qy    1407 cggcccagcccagcag 1423
           ||||||||||||||||
Db   20332 CGGCCCAGCCCAGCAG 20316

Query Match       0.7%;  Score 17;  DB 3;  Length 68750;
Best Local Similarity 100.0%;  Pred. No. 85;
Matches   17;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Search completed: October 28, 2001, 21:03:46
Job time: 5450 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:            October 28, 2001, 18:05:01 ; Search time 1834.94 Seconds
                                              (without alignments)
                                              12781.088 Million cell updates/sec

Title:             US-09-434-382-1
Perfect score:     2481
Sequence:          1 atgtgggcgcttgctcgct.........agaaggtcaagagccagtga 2481

Scoring table:     OLIGO_NUC
                   Gapop 60.0 , Gapext 60.0

Searched:          10228115 seqs, 472642675Ö residues

Word size :        8

Total number of hits satisfying chosen parameters:       20358679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :         EST:*

1:   gb_est1:*
2:   gb_est2:*
3:   gb_est3:*
4:   gb_est4:*
5:   gb_est5:*
6:   gb_est6:*
7:   gb_est7:*
8:   gb_est8:*
9:   gb_est9:*
10:  gb_est10:*
11:  gb_est11:*
12:  gb_est12:*
13:  gb_est13:*
14:  gb_est14:*
15:  gb_est15:*
16:  gb_est16:*
17:  gb_est17:*
18:  gb_est18:*
19:  gb_est19:*
20:  gb_est20:*
21:  gb_est21:*
22:  gb_est22:*
23:  gb_est23:*
24:  gb_est24:*
25:  gb_est33:*
26:  gb_est34:*
27:  gb_est35:*
28:  gb_est36:*
29:  gb_est37:*
30:  gb_est38:*
31:  gb_est39:*
32:  gb_est40:*
33:  em_estba:*
34:  em_estfun:*
35:  em_esthum1:*
36:  em_esthum2:*
37:  em_esthum3:*
38:  em_esthum4:*
39:  em_esthum5:*
40:  em_esthum6:*
41:  em_esthum7:*
42:  em_esthum8:*
43:  em_esthum9:*
44:  em_esthum10:*
45:  em_esthum11:*
46:  em_esthum12:*
47:  em_esthum13:*
48:  em_esthum14:*
49:  em_esthum15:*
50:  em_esthum16:*
51:  em_esthum17:*
52:  em_esthum18:*
53:  em_esthum19:*
54:  em_esthum20:*
55:  em_esthum21:*
56:  em_esthum22:*
57:  em_esthum23:*
58:  em_esthum24:*
59:  em_esthum25:*
60:  em_esthum26:*
61:  em_esthum27:*
62:  em_esthum28:*
63:  em_estin1:*
64:  em_estin2:*
65:  em_estin3:*
66:  em_estin4:*
67:  em_estin5:*
68:  em_estom1:*
69:  em_estom2:*
70:  em_estov1:*
71:  em_estov2:*
72:  em_estpl1:*
73:  em_estpl2:*
74:  em_estpl3:*
75:  em_estpl4:*
76:  em_estpl5:*
77:  em_estpl6:*
78:  em_estpl7:*
79:  em_estpl8:*
80:  em_estpl9:*
81:  em_estpl10:*
82:  em_estro1:*
83:  em_estro2:*
84:  em_estro3:*
85:  em_estro4:*
86:  em_estro5:*
87:  em_estro6:*
88:  em_estro7:*
89:  em_estro8:*
90:  em_estro9:*
91:  em_estro10:*
92:  em_estro11:*
93:  em_estro12:*
94:  em_estro13:*
95:  em_estro14:*
96:  em_estro15:*
97:  em_estro16:*
98:  em_estro17:*
99:  em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estp1l1:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*

190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_Other:*
207: em_gss_Pln1:*
208: em_gss_Pln2:*
209: em_gss_Pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
RESULT  1
LOCUS       BE260495      676 bp   mRNA      EST  26-OCT-2000
DEFINITION  6011S0702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
            mRNA sequence.
ACCESSION   BE260495
VERSION     BE260495.1  GI:9131807
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            High quality sequence stop: 672.
            Plate: LLCM176 row: d column: 01
FEATURES             Location/Qualifiers
     source          1..676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503184"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT    154 a    207 c    176 g    139 t
ORIGIN
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	676	27.2	676	165	BE260495	601150702
2	639	25.8	712	107	AU126037	601150702
3	609	24.5	666	108	AU141334	AU126037
4	601	24.2	762	172	AU124671	AU141334
5	597	24.1	944	172	BF969043	AU124671
6	590	23.8	664	166	BE383336	BF969043
7	590	23.8	692	166	BE382353	601298249
8	561	22.6	920	153	BG386090	601298656
9	553	22.3	553	148	BF434169	7099r04
10	543	21.9	827	153	BG396395	7099r04.x
11	530	21.4	614	152	BG335963	602459323
12	528	21.3	761	140	BE795820	602404708
13	526	21.2	745	145	BF196304	601509856
14	525	21.2	696	108	BE794311	7n6a10.x
15	521	21.0	688	140	BF797306	601591442
16	519	20.9	960	169	AU143668	602256926
17	507	20.4	670	108	BF057321	AU143668
18	506	20.0	506	143	BF794577	7k19c01.x
19	497	20.0	953	169	BG519751	602558843
20	491	19.8	884	154	BG386348	602578902
21	491	19.8	1012	153	AU138795	602455550
22	478	19.3	657	108	BF311926	AU138795
23	477	19.2	915	147	BE619259	601897767
24	475	19.1	823	138	BG471513	601473130
25	465	18.7	531	154	AA243700	602513052
26	458	18.5	611	17	BG471513	zr68g08.s
27	453	18.3	451	141	AI200296	qf86b12.x
28	451	18.2	812	148	BE867512	601443010
29	451	18.1	938	165	BF433563	7q55007.x
30	450	18.1	938	118	BE260626	601146116
31	447	18.0	493	166	AW575677	UI-HF-BM0
32	447	18.0	612	7	BE304720	601106236
33	442	17.8	493	20	AA455121	zx78c04.s
34	438	17.7	421	113	AI468143	tf92g05.x
35	436	17.6	536	152	AW206103	UI-H-BI1-
36	421	17.0	452	147	BG327066	602426274
37	421	17.0	457	114	BF309436	601892128
38	406	16.4	474	4	AW304130	xs13e05.x
39	405	16.3	947	150	AA291670	zt37d04.s
40	404	16.3	992	159	BF525432	602069517
41	404	16.2	627	167	BE747163	601577254
42	402	16.2	627	167	BE386924	601274815
43	400	16.1	574	165	BE250309	600943455
44	400	16.1	397	168	BF732987	nae07b01.
45	387	15.6	878	147	BF338283	602035507

ALIGNMENTS

```
Query Match      27.2%;  Score 676;  DB 165;  Length 676;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches  676;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  1093  accagcacttggttcctgaatgagagaactgtgtcccagttcacaaacttcgcagcacaag  1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  accagcacttggttcctgaatgagagaactgtgtcccagttcacaaacttcgcagcacaag  60

Qy  1153  attcaaaacccagctcaacctcatccaccccggacaatcctcccccctgctcaacagtttccgc  1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  attcaaaacccagctcaacctcatccaccccggacaatcctcccccctgctcaacagtttccgc  120

Qy  1213  tgtaagaaggagggcccccaccccagtgttgccatgttcagggtgaatgcctcctcaag  1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  tgtaagaaggagggcccccaccccagtgttgccatgttcagggtgaatgcctcctcaag  180

Qy  1273  taccagctccgtccaggagggagtggcagaggatggccattattacttgcaatcctgag  1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  taccagctccgtccaggagggagtggcagaggatggccattattacttgcaatcctgag  240

Qy  1333  gaattcatagttgagcgcgctgcagctccccaacttccagcagcaggcgtcaaggagtacagg  1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  gaattcatagttgagcgcgctgcagctccccaacttccagcagcaggcgtcaaggagtacagg  300

Qy  1393  aggagtcgcgcaggacgcccagcccagccagagaaagaagtcagtacccagaaaatcatc  1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  aggagtcgcgcaggacgcccagcccagccagagaaagaagtcagtacccagaaaatcatc  360

Qy  1453  ttccttggaacaggtgtctgccatcccgatgaagatcgaaatgtcagtgcacacttgtc  1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  ttccttggaacaggtgtctgccatcccgatgaagatcgaaatgtcagtgcacacttgtc  420

Qy  1513  aacataagccccgaccacgtctctgtactggactgtgtggaggacaattgggacgctg  1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  aacataagccccgaccacgtctctgtactggactgtgtggaggacaattgggacgctg  480

Qy  1573  tgccgtcattacgagacccaggtggacaaggtcctggcaccctggctgctgtgttgtg  1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  tgccgtcattacgagacccaggtggacaaggtcctggcaccctggctgctgtgttgtg  540

Qy  1633  tcccacctgcacgcagcagatcaccacacgggcttgccaagtatcttgctgcagagagaacgc  1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  tcccacctgcacgcagcagatcaccacacgggcttgccaagtatcttgctgcagagagaacgc  600
```

```
Qy     1693 gccttggcatcttttgggaaagccgcttcaacccttttggtttggttgccccaaccaagtc 1752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GCCTTGGCCATCTTTTGGGAAAGCCGCTTCACCCTTTTGGCTTGGTTGGTTGCCCCAACCAGTC 660

Qy     1753 aaagccttggctccagc 1768
            |||||||||||||||||
Db      661 AAAGCCTTGGCTCCAGC 676
```

RESULT 2

```
LOCUS        AU126037      712 bp      mRNA                           EST      23-OCT-2000
DEFINITION   AU126037 NT2RM4 Homo sapiens cDNA clone NT2RM4002610 5', mRNA
             sequence.
ACCESSION    AU126037
VERSION      AU126037.1  GI:10950753
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 712)
  AUTHORS    Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
             Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
             Isogai,T.
  TITLE      HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
             Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
             ,Y., Sugano,S., Isogai,T.)
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Takao Isogai
             Genomics Laboratory
             Helix Research Institute
             1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             Tel: 81-438-52-3951
             Fax: 81-438-52-3952
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
             Research Institute; cDNA library construction: Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RM4002610"
                     /clone_lib="NT2RM4"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
BASE COUNT      173 a    198 c    204 g    134 t    3 others
ORIGIN
```

Query Match   25.8%;   Score 639;   DB 107;   Length 712;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 639;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy        1 atgtgggcgctttgctctcgctcgtgcgtgcgtcccggccggacgacgcaccatgtcgcaggggacgc 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       22 ATGTGGGCGCTTTGCTCGCTCGTCCTGCGCGTCCCGGCCGGACGACGCACCATGTCGCAGGGGACGC 81

Qy       61 accatatcgcaggcaccgccccgccgccccgcgagcggcgcgcaaggacccgctcggcaacctg 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       82 ACCATATCGCAGGCACCGCCCCGCCGCCCCGCGAGCGGCGCGCAAGGACCCGCTCGGCAACCTG 141

Qy      121 cgcaccgagagaagccggaccgtcagggtgtctcagggacccccaaacaccgtgtaactg 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      142 CGCACCGAGAGAAGCCGGACCGTCAGGGTGTCTCAGGGACCCCCAAACACCGTGTACCTG 201

Qy      181 caggtcggtcggcagcgggtagcgggcgcgccggactcggcgggtgttcctccttcgagttc 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      202 CAGGTCGGTGGCAGCGGGTAGCGGGCGCGCCGGACTCGGCGGGTGTTCCTCCTTCCGAGTTC 261
```

RESULT 3

```
LOCUS        AU141334      666 bp      mRNA                           EST      25-OCT-2000
DEFINITION   AU141334 THYRO1 Homo sapiens cDNA clone THYRO1000421 5', mRNA
             sequence.
ACCESSION    AU141334
VERSION      AU141334.1  GI:11002855
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 666)
  AUTHORS    Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
             Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
             Isogai,T.
  TITLE      HRI human cDNA project
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Takao Isogai
             Genomics Laboratory
             Helix Research Institute
             1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             Tel: 81-438-52-3951
             Fax: 81-438-52-3952
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
             Research Institute; cDNA library construction: Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="THYRO1000421"
                     /clone_lib="THYRO1"
                     /tissue_type="thyroid gland"
                     /note="Vector: pME18SFL3"
BASE COUNT      158 a    186 c    190 g    129 t    3 others
ORIGIN
```

Query Match   24.5%;   Score 609;   DB 108;   Length 666;
Best Local Similarity   100.0%;   Pred. No. 0;

```
Qy      241 aaccgtgtatctctcaacttgtggagaagggcgttcagagagactcatgtcagaggacacaagtta 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      262 AACCGGTATCTCTCTCAACTGTGGAGAAGGGCGTTCAGAGACTCATGCAGAGCACAAAGTTA 321

Qy      301 aaggttgctcgcctggacacatatttcctgacacgaatgcactggtctaatgtgtggggc 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      322 AAGGTTGCTCGCCTGGACACATATTTCCTGACAACATGATTCTCGACACGAATGCACTGGTCTAATGTTGGGGGC 381

Qy      361 ttaagtggaatggattcttacttttaaaggaaaaccgggcttccaaaagtgtgttacttctgga 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      382 TTAAGTGGAATGGATTCTTACTTACTTTAAAGGAAAAACCGGGCTTCCAAAAGTGTGTTACTTTCTGGA 441

Qy      421 cctccacaactgggaaaatatacctcgaagcaatcaaaaatatttctggttcattgaaagga 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      442 CCTCCACAACTGGGAAAAATACCTCGAAGCAAATCAAAAATATTTCTGGTTCATTGAAAGGA 501

Qy      481 atagaactgggtgtgcggcccccactcctgccccagaatacgggggatgaaacatgacagtt 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      502 ATAGAACTGGGTGTGCGGCCCCCACTCTGCCCCAGAATACGAGGGATGAAAACATGACAGTT 561

Qy      541 taccagatcccccatacacagtgaacacagagagagagagggggaaaaagcacacaaccatggacaagtcca 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      562 TACCAGATCCCCATACACAGTGAACACAGTGAACAGAGAGGGGAAAGCACACCATGGCACAAGTCCA 621

Qy      601 gaaaggcgcctcagcaggtcagtccagagagcgatcttca 639
            ||||||||||||||||||||||||||||||||||||||||
Db      622 GAAAGGCGCCTCTCAGCAGGTCAGTCCAGACGGATCTTCA 660
```

RESULT 4

Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgtgggcgcttgctcgctcgcgtcgcgtgcgggacggcaccatgtcgcaggagacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     29 ATGTGGGCGCTTTGCTCGCTCGCGTGCGGGACGGCACCATGTCGCAGGAGACGC 88

Qy     61 acacatatcgcagtcaccgccgcccgccgaagcggccgcaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89 ACCAATATCGCAGGCACCCCGCCGCGCAAGGCGCGGCCAAGGACCCGCTGCGGCACCTG 148

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgtcctccgggtgcccaaaacacctgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGTCTCCGGGTGCCCAAAACACCTGTACCTG 208

Qy    181 caggtggttgcaggcgtagcccgggaactcgggcgccgcgcctctacgtcttcctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 CAGGTGGTTGCAGGCGGGTAGCCCGGGGACTCGGGCGCCGCGCCTCTACGTCTTCCTACCTG 268

Qy    241 aaccggtatctcttcaactgtggaagaaggcgttcagagaactcatgtgagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCGAGGACACAAGTTA 328

Qy    301 aaggttgctccgcctggaacaacatattcctggacacgaatgcacctggtctaatgtgtggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 AAGGTTGCTCCGCCTGGACACAACATATTCCTGGACACGAATGCACTGGTCTAATGTTGGGGGC 388

Qy    361 ttaagtggaatgattcttacttaaagagaaaaccgggcttccaaagttgtgtacttctggga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    389 TTAAGTGGAATGATTCTTACTTTAAAGGAAAACCGGGCTTCCAAAGTGTGTGTACTTTCTGGA 448

Qy    421 ccctccacaacctgaaaaataccctcgaaagcaaatcaaaaatatttttctgtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 CCTCCACAACTGGAAAAAATACCCTCGAAAGCAATCAAAAAATATTTTTCTGGTTCATTGAAAGGA 508

Qy    481 atagaactggctgtgcggcgcccaactctgccccagaatacgaggatgaaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509 ATAGAACTGGCTGTGTGGCGCCCCACTCTGCCCCAGAATACGAGGATGAAAACCATGACAGTT 568

Qy    541 taccagatccccataacaacagtgaacagaggagggagggaaagaggaacacatggcagctcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    569 TACCAGAATCCCCATACACAAGTGAACAGAGGAGGAGGGAAAAGCACACCATGGCAGAGTCCA 628

Qy    601 gaaaggcct 609
          |||||||||
Db    629 GAAAGGCCT 637
```

FEATURES   Location/Qualifiers
   source   1..762
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="NT2RM4000375"
                   /clone_lib="NT2RM4"
                   /cell_type="teratocarcinoma"
                   /cell_line="NT2"
                   /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                   precursor cells"

BASE COUNT   180 a   212 c   165 t   3 others

ORIGIN

RESULT 5

Query Match 24.2%; Score 601; DB 107; Length 762;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 651; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    852 cacctcatgaagagaagaagaattttgggctgaagaagctgtgtactcctccagatcctgtgc 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACTCATGAAGGAAGACAGAGATTTTGGCGTGAACAGCGTGTACTCCTCCAGATCCTGTGC 60

Qy    912 tgcttttgtggtggtagaatgtccagatgaaagcttcaatcaaacccatctgtgagaatgc 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTTTTGTGGTGGTAGAATGTCCAGAATGAAAGCTTCAATCAAACCCATCTGTGAGAATGC 120

Qy    972 cacccttcaagagggtaccacaaggacaaggcagatgccccgtgccttggtggttcaacatggc 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACCCTTTCAGAGGGTACCACAAGGAAAGGCAGATGCCCCGTGGGCCTTGGTGGTTCACATGGC 180

Qy   1032 cccagcatcctgctctgtgtgtggacagcagcaagtgaccagcagtggagaggtttgggctga 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCCAGCATCCTGGTGCTTGTGGCAGCAGCAAGTGACCAGCAGTGGAGAGGTTTGGGCCTGA 240

Qy   1092 caccagcacttggtcctgaatgagaactgtgcctcaagttcaacaacttcgcagcacaca 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCAACAACTTCGGCAGCCACA 300

Qy   1152 gattcaaaaccagctcaaacctcatccacacccggacacctccccctgctcaccagtttccg 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GATTCAAAACCAGCTCAAACCTCATCCACACCCGGACACATCTTCCCCCTGCTCACCAGTTTCCG 360

Qy   1212 ctgtaagaaggaggggccccacctcagtgtgcccatgttcaggggtgaaatgctcctcaa 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGTAAGAAGGAGGGGCCCCACCTCAGTGTGCCCATGTTCAGGGGTGAAATGCCTCTCAA 420

Qy   1272 gtaccagctccgtccgtcccaggaggggagtggcagaggagatgccattattacttgcaatcctga 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTACCAGCTCCGTCCCAGGAGGGGAGTGGCAGGATGCCATTATTACTTGCAATCCTGA 480

Qy   1332 ggaattccatagttgaggccgctgcagcttcccaacttccagagacgagacgtgcaggagtacag 1391
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GGAATTCATAGTTGAGGCCGCTGCAGCTTCCCAACTTCCATCAGAGCGTGCAGGAGTACAG 540

Qy   1392 gaggagtgcgcaggacggccccagcgccagcagaagaaagcagtacccgagaaatcat 1451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAAGAAAGAAGTCAGTACCCAGAAATCAT 600

Qy   1452 cttcccttggaacaggtctgccatccgatgaagaagattcgaaaatgtcagtgcc 1503
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTTCCTTGGAACAGGGTCTGCCATCCGATGAAGAATTCGAAAATGTCAGTGCC 652
```

ACCESSION   BF969043
VERSION   BF969043.1 GI:12336258
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 944)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Plate: LLAM9995 row: d column: 18
              High quality sequence stop: 591.
FEATURES   Location/Qualifiers
     source   1..944
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4357985"
                     /clone_lib="NIH_MGC_84"
                     /tissue_type="adrenal cortex carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
                     primed. Average insert size 1.229 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
BASE COUNT   200 a   266 c   264 g   214 t
ORIGIN

Query Match   24.18; Score 597; DB 172; Length 944;
Best Local Similarity   100.0%;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   796 gttgggacagctgccatcgctcccatcattgtgctgctgtcaaggacggggaaaagcatact 855
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTTGGGACAGCTGCCATCGCTCCCATCATTGTGCTGCTGTCAAGGACGGGGAAAAGCATACT 60

Qy   856 catgaagagaaagaagagagtttggctgaagagctgtgtgtactcctccagatcctggtgctgct 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATGAAGAGAAAGAAGAGAGATTTTGGCTGAAGAGCGTGTGTACTCCTCCAGATCCTGGTGCTGCT 120

Qy   916 tttggtggttagaagtgtccagatgaaaagcttcaattcaacccatctgtgagaaatgcaac 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TTTGGTGGTTAGAAGTGTCCAGATGAAAAGCTTCATTCAACCCATCTGTGAGAATGCCACC 180

Qy   976 tttcagaggtaccaaaggaaaggcagatgccccccgtggcctggggttcacatggcccca 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTTCAGAGGTACCAAAGGAAAGGCAGATGCCCCCGTGGCCTGGGGTTCACATGGCCCCA 240

Qy   1036 gcatctgtgctgtgagcagcaggtaccagcagttgatgaggaggtttgggcctgacacc 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCATCTGTGCTGTGTGGACAGCAGGTACCAGCAGTTGGGACGAGAGGTTTGGGCCTGACACC 300

Qy   1096 cagcacttggtccctgaatgagaacttgtgcctcacacaagatt 1155
          ||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGCACTTGGTCCCTGAATGAGAACTTGTGCCTCACACAAGATT 360

Qy   1156 caaaccaagctcaaactcatccacccggacatcttccacccagttccgctgt 1215
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAAACCCAAGCTCAAACTCATCCACCCGGACATCTTCCCCCAGTTTCCGCTGT 420

Qy   1216 aagaaggagggcccacccctcagtgtgcccatggttcaaggtgaatgcgctcctcaagtac 1275
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT   6

LOCUS   BE383336   664 bp   mRNA   EST   21-JUL-2000
DEFINITION 601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
            mRNA sequence.
ACCESSION   BE383336
VERSION   BE383336.1 GI:9287701
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 664)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM313 row: a column: 13
            High quality sequence stop: 662.
FEATURES   Location/Qualifiers
     source   1..664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3628308"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOT87; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   154 a   199 c   174 g   137 t
ORIGIN

Query Match   23.8%; Score 590; DB 166; Length 664;
Best Local Similarity   100.0%;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db   421 AAGAAGGAGGGCCCACCCCTCAGTGTGCCCATGGTTCAAGGTGAATGCGCTCCTCAAGTAC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1276 cagctccgtcccacccaggagggagtggcagaggaggggatgccaattacttgcaatcctgaggaa 1335

Db   481 CAGCTCCGTCCCAGGGAGGGAGTGGCAGGAGGGGATGCCATTATTACTTGCAATCCTGAGGAA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1336 ttcataagttgagggcgctgcagcttcccaacttccagcaagagcgtgcaaggaagtacaag 1392

Db   541 TTCATAAGTTGAGGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAAGGAAGTACAGG 597
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1093 acccagcacttggtccctgaatgagaacttgtgcctcagttcacaacctttcgcagccacag 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACCCAGCACTTGGTCCCTGAATGAGAACTTGTGCCTCAGTTCACAACCTTCGCAGCCACAAG 60

Qy   1153 attcaaaccaagctcaaactcatccacccggacatctccccctgccctcaccagttccgc 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATTCAAACCCAGCTCAAACTCATCCACCCGGACATCTCCCCCTGCCCTCACCAGTTTCGC 120

Qy   1213 tgtaagaagaggaggggcccacccctcagtgtgccccatggttcaaggtgaatgcgctcctcaag 1272
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGTAAGAAGAGGAGGGGCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAG 180

RESULT 7

```
Qy   1273 tacaagctccgtccaggagggagtggcaggaggaatgcattattacttgcaatcctgag 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TACCAGCTTCCGTCCCAGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAG 240

Qy   1333 gaattcataggttgaggcgtcagcttcccaacttcccagacagaggagtacagg 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAATTCATAGTTGAGGCGTCAGCTTCCCAACTTCCACAGAGACGTGCAGGAGTACAGG 300

Qy   1393 aggagtgcgcaggacggacggccagcccagcaggaatcatc 1452
          |||||||||||||||||||||||||||||||||||||||||
Db    301 AGGAGTGCGCGAGGACGGACGGCCCAGCCCAGCAGGAAATGAGAGAGAGTCACGTACCCAGAAAATCATC 360

Qy   1453 ttccttggaacaagggtctgccaatcccgatgaagattcgaaatgtgccacacttgtc 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTC 420

Qy   1513 aacataagcccgacacagtctctgctactggacgtggtggagggcacatttggcaagctg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACATAAGCCCGACACAGTCTCTGCTACTGGACGTGGTGGAGGGCACATTTGGGCAAGCTG 480

Qy   1573 tgcgtcattacggagaccaaggtggacaggggtgtcctgggacgtgtgggcctggtgttgg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCGTCATTACGGAGACCAAGGTGGACAGGGGTGTCCTGGGCACCTGGTCTGTGTTTGTG 540

Qy   1633 tcccacctgcacgcagatcaccacaacggcgcttgccaagtatcttgctgca 1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCCACCTGCACGCAGATCACCACAACGGGGCTTGCCAAGTATCTTGCTGCA 590
```

LOCUS   BE382353   692 bp   mRNA   EST   21-JUL-2000
DEFINITION 601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
            mRNA sequence.
ACCESSION   BE382353
VERSION   BE382353.1 GI:9327718
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 692)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM314 row: o column: 13
            High quality sequence stop: 600.
FEATURES   Location/Qualifiers
     source   1..692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629028"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /note="Organ: brain; Vector: pOTB7;
                     lab_host="DH10B (phage-resistant)"
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   152 a   193 c   197 g   150 t
ORIGIN

Query Match   23.8%;   Score 590; DB 166; Length 692;
Best Local Similarity 100.0%;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1475 tccgatgaagattcgaaatgtgccacacttgtcaaacataagcccgacacagtctctc 1534
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCCGGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCGACACGTCTC 60

Qy   1535 tgctactggacgtggtggagggcacatttggcaagctgtgcgtcattacggag 1594
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTACTGGACGTGGTGGAGGGCACATTTGGGCAAGCTGTGCGTCATTACGGAGACCAAGG 120

Qy   1595 tggacaggggtcctgggcaccctggctgtgtttgtgtcccacctgcacgcagatcacc 1654
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGACAGGGGTCCTGGGCACCCTGGTCTGTGTTTGTGTCCCACCTGCACGCAGATCACC 180

Qy   1655 acacgggcttgccaagtatcttgctgcagagagaacgcgccttggcatcttgggaaagc 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACACGGGGCTTGCCAAGTATCTTGCTGCAGAGAGGAACGCGCCTTGGCATCTTTGGGAAAGC 240

Qy   1715 cgcttcaccctttgctgggtgggttgccccaaccaagctcaaaagcctggctccagcagtacc 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGCTTCACCCTTTGCTGGGTGGGTTGCCCCAACCAAGCTCAAAAGCCTGGCTCCAGCAGTACC 300

Qy   1775 acaaccagtgccaggaggtcctgccacacacaagttatgaagttccctgccaaatgccttcagg 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACAACCAGTGCCAGGAGGTCCTGCCACACACAAGTATGAAGTTCCCTGCCAAATGCCTTCAGG 360

Qy   1835 aaggggctgagaatcccagtcctgcagtgtgaaagattcggaacatcgacaat 1894
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGGGCTGAGAATCCCAGTCCTGCAGTGTGAAAGATTGATCAGTTGCGTGTTGCGAACAT 420

Qy   1895 gtgattggaagagttcagacctgtctggttgcggcagctgcaagcatgcgtttggctgtg 1954
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTGATTTGGAAGAGTTTCAGACCTGTGTCTGGTGCGGCAGCACTGCAAGCATGCGTTTGGCTGTG 480

Qy   1955 cgctggtgcacacctctggctggtcatattccgggggacacacaatgccctgcgagg 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGCTGGTGCACACCTCTGGCTGGAAGTGGGTCTTATTCCGGGGACACACATGCCCCTGCGAGG 540

Qy   2015 ctctgtccggatgggaagatgccacccttcctgatacatgaagccacc 2064
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCTGGTTCCGGATGGGAAGATGCCACACCTTCCTGATACATGAAGCCACC 590
```

RESULT 8

LOCUS   BG386090   920 bp   mRNA   EST   12-MAR-2001
DEFINITION 6024552641F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583679 5',
            mRNA sequence.
ACCESSION   BG386090
VERSION   BG386090.1 GI:13279536
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 920)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1307 row: p column: 16

FEATURES
source
1..920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583679"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) and Superscript II RT (life Technologies)"

High quality sequence stop: 736.
Location/Qualifiers

BASE COUNT   196 a   261 c   272 g   190 t
ORIGIN

Query Match   22.6%;   Score 561;   DB 153;   Length 920;
Best Local Similarity   99.9%;   Pred. No. 4.6e-285;
Matches 681; Conservative   0; Mismatches   0;   Indels   1;   Gaps   1;

Qy   916 tttgtgtggtagaatgtccagatgaaagcttcattcaaccatctgtgagaatgcaacc 975
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 TTTGTGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCATCTGTGAGAATGCCACC 80

Qy   976 tttcagaggtaccaaggaaagcgagatgcccgtgccctggtggttcacatggcccca 1035
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 TTTCAGAGGTACCAAGGAAAAGCGAGAATGCCCCGTGGCCTTGGTGGTTCACATGGCCCA 140

Qy   1036 gcatctgtcgttgtggacagcaggtaccagtggatggagaagtttgggcctgaacc 1095
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 GCATCTGTCGTTGCTTGTGGACAGCAGGTACCAGTGGATGGAGAAGTTTGGGCCTGAACAC 200

Qy   1096 cagcacttggtcctgaatgagaactgtgcctcagttcacaaacacttcgcagcacaagatt 1155
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 CAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAAACCTTCGCAGCACAAGATT 260

Qy   1156 caaacccaagctcaaacctcatcaccccgaacatcttccccctgctcaccagttccgctgt 1215
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 CAAACCCCAGCTCAAACCTCATCACCCCGACATCTTCCCCGTCACCAGTTCCGCTGT 320

Qy   1216 aagaagaaggagggcccaccacctcagtgtgccccatggttcaaggggtgaaatgcctcctcaagtac 1275
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 AAGAAGAAGGAGGGCCCCACCACCTCAGTGTGCCCCATGGTTCAAGGGTGAAATGCCTCCTCAAGTAC 380

Qy   1276 cagctccgtccaaggaggagtggccagaggggatgccattattacttgtaatcctgaggaa 1335
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 CAGCTCCGTCCCAAGGAGGAGTGGCCAGAGGGGATGCCATTATTACTTGTAATCCTGAGGAA 440

Qy   1336 ttcatagttgaggcgttgcagcttcccaacttccagcaagagcggtgcaggagtacaggagg 1395
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   441 TTCATAGTTGAGGCGTTGCAGCTTCCCAACTTCCAGCAAGAGCGGTGCAGGAGTACAGGAGG 500

Qy   1396 agtgcgcagacgagagccccagcccagcagagaaagaaagtcagtaccccagaaatcatcttc 1455
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   501 AGTGCGCAGACGGAGACGCCCCAGCCCAGCAGAGAAAGAAAGTCAGTACCCCAGAAATCATCTTC 560

Qy   1456 cttggaacaaggtctgccatccggatgaaagtgtcagtgtgccacacttgtcaac 1515
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   561 CTTGGAACAAGGTCTG-CATCCGGATGAAAGATGTCAGTGTGCCACACTTGTCAAC 619

Qy   1516 ataagccccgacaacgtctctgctcctggaactggacttgttgaggaggcacacattggaacgctgtgc 1575
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   620 ATAAGCCCGACAACGTCTCTGCTCCTGGAACTGGACTTGTTGAGGGCACATTTGGGACTGTGC 679

Qy   1576 cgtcattacggaagaccaaggtgg 1597
            ||||||||||||||||||||||||
Db   680 CGTCATTACGGAGAGACCAAGGTGG 701

RESULT   9

LOCUS   BF434169   553 bp   mRNA   EST   29-NOV-2000
DEFINITION   7o99f04.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644670 3'
    similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF434169
VERSION   BF434169.1   GI:11464441
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 553)
AUTHORS   Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
    R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
    I.M.A.G.E. Consortium DNA Sequencing by: Washington University
    Genome Sequencing Center
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    This clone distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL, send email to:
    info@image.llnl.gov
    Seq primer: -40UP from Gibco
    Seq primer: -40UP from Gibco
    High quality sequence stop: 507.

FEATURES
source
1..553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3644670"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGGAGGCGGCGGCCGCGGGACAATTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
Location/Qualifiers

BASE COUNT   143 a   150 c   153 g   107 t
ORIGIN

Query Match   22.3%;   Score 553;   DB 148;   Length 553;
Best Local Similarity   100.0%;   Pred. No. 7.7e-281;
Matches 553; Conservative   0; Mismatches   0;   Indels   0;   Gaps   0;

Qy   91 cggcgcgcgcaaggagaccccgcgtgcgcacctgcgcacgcgagagaagcgccgagaccgtccgggg 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CGGCGCGCGCAAGGAGACCCCGCGTGCGCACCTGCGCACGCGAGAGAAGCGCCGAGACCGTCCGGGG 60

Qy   151 tgctccggcgcgcccaaacaccgtgtactgcagtgtcagttgcagggggtagcgggactcg 210
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGCTCCGGCGCGCCCAAACACCGTGTACCTGCAGTGGTGGTGGCAGGGGGTAGCCGGGACTCG 120

Qy   211 ggcgccgcccgctacgtcttcccggagtcaaccggtatctctcaaactgtggagaagc 270
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCGCCGCGCTCTACGTCTTCTCCGGAGTTCAACCGGTATCTCTTCAAACTGTGGAGAAGC 180

Qy   271 gttcagagactcatgcaggagagcacaagttaaagtttgctcgcctggaacacatattcctg 330
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTTCAGAGACTCATGCAGGAGAGCACAAGTTAAAGGTTGCTCGCCCTGGACAACACATATTCCTG 240

Qy   331 acacgaatgcaactgtctaatgttgggggcttaagtggaatgagattcttacttaaaggaa 390
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Qy    241  ACACGAATGCACTGTGTCTAATGTGGGGCGGCTTAAGTGGAATGAATTCTTACTTAAAGGAA  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391  accgggcttccaaagtggtagtacttcctggaccttccacacctggaaaaataacctcgaagca  450

Qy    301  ACCGGGCTTCCAAAGTGTGTNACTTTCTGGACCTTCCACACTGGAAAAATACCTCGAAGCA  360
           |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db    451  atcaaaaatattttctggtccattgaaaatgaaataaaactggtctggtgcggcccactctgcc  510

Qy    361  ATCAAAAATATTTTCTGGTCCATTGAAAGGAATAGAGAACTGGCTGTCGGGCCCCACTCTGCC  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    511  agggagaagcaccaaccatggcagagtccagaaaggccctcctcagcgcaggcctcagtccagag  570

Qy    421  CCAGAAATACGAGGATGAAAACCATAGACAGTTACCACAGTTGACAGAGG  480
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    571  ccagaataccgaagtgaaaccatcccacataacacagtgaacagag  630

Qy    481  AGGGGAAAGGCACCAACCATGGCAGAGTCCAGAAAGGCCCTCCTCAGCGCAGGCCTCAGTCCAGAG  540
           ||||||||||||||||||||||||||||||||||||||||||||
Db    631  cgatctttcagact  643

Qy    541  CGATCTTTCAGACT  553
```

RESULT 10

LOCUS   BG396395   827 bp   mRNA   EST   12-MAR-2001
DEFINITION   6024593371 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581700 5',
             mRNA sequence.
ACCESSION   BG396395
VERSION   BG396395.1 GI:13289941
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 827)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM1302 row: n column: 05
             High quality sequence stop: 779.
FEATURES   Location/Qualifiers
     source   1..827
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4581700"
                     /clone_lib="NIH_MGC_16"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   205 a   217 c   226 g   179 t
ORIGIN

Query Match   21.9%;   Score 543; DB 153;   Length 827;
Best Local Similarity 100.0%; Pred. No. 1.5e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    707  actcttccctgtgtcgttagtcttcattcgtgtctgaagcttcacttaaagagaagaaacttcttgg  766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22  ACTCTTCCCTGTGTCGTTAGTCTTCATCGTGTCTGAAGCTTCACTTAAAGAGAAGAAACTTCTTGG   81

Qy    767  tgctcaaagcaaagaggaggatgggcctcccagttgggacagctgccatcgctcccaatcatg  826
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82  TGCTCAAAGCAAAGAGGAGGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCAATCATTG  141

Qy    827  ctgctgtcaaggacgggaaaagcatcactcatgaagagaagaagaagaaagaagaagagc  886
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142  CTGCTGTCAAGGACGGGAAAAGCATCACTCATGAAGAGAAGAGACGAGATTTTGGCTGAAGAGAGC  201

Qy    887  tgtgtactctcccacgatcctgtgtgctgcttttgttggtagaaatgttcaagataaagct  946
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202  TGTGTACTCTCCCACGATCCTGTGCTGCTTTTTGTTGGTGGTAGAATGTTCCAGATGAAAGCT  261

Qy    947  tcattcaacccatctgtgaagaatgccaacctttcagaggggtaccaaggaaaggcagatgcc  1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262  TCATTCAACCCATCTGTGAAGAATGCCAACCTTTCAGAGGGGTACCAAGGAAAGGCAGATGCC  321

Qy   1007  ccgtgggccttggtgttcacatggcccccagacatctcgtctgctcgtgtgtgacgaactgtgctt  1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322  CCGTGGGCCTTGGTGTTCACATGGGCCCCCAGACATCTCGTCTGCTCGTGTGTGACGAACTGTGCTT  381

Qy   1067  agtggatggagagtttgggcctgacacccgacactggtcctgaatgagagaacttgtgcct  1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382  AGTGGATGGAGAGTTTGGGCCTGACACCCGACACTGGTCCTGAATGAGAGAACTTGTGCCT  441

Qy   1127  cagttcacaaccttcgcacacaaagaattcaaacccagctcaaacctcatcaccccggaca  1186
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442  CAGTTCACAACCTTCGCACACAAAGAATTCAAACCCAGCTCAAACCTCATCACCACCGGACA  501

Qy   1187  tcttccccctgtctcaccacagtttccgctgtaagaaaggaggggccccacccctcagttgtgccca  1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502  TCTTCCCCCTGTCTCACCACAGTTTCCGCTGTAAGAAAGGAGGGGCCCCACCCCTCAGTTGTGCCCA  561

Qy   1247  tgg  1249
           |||
Db    562  TGG  564
```

RESULT 11

LOCUS   BG335963   614 bp   mRNA   EST   27-FEB-2001
DEFINITION   6024704708F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542553 5',
             mRNA sequence.
ACCESSION   BG335963
VERSION   BG335963.1 GI:13142401
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 614)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1222 row: o column: 02
             High quality sequence stop: 614.
FEATURES   Location/Qualifiers
     source   1..614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4542553"

COMMENT   Contact: Robert Strausberg, Ph.D.
JOURNAL   Unpublished (1999)
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
REFERENCE   1 (bases 1 to 761)
                 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Homo sapiens
SOURCE   human.
KEYWORDS   EST.
VERSION   BE795820.1 GI:10217018
ACCESSION   BE795820
                 mRNA sequence.
DEFINITION   601590856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
LOCUS   BE795820   761 bp   mRNA   EST   20-SEP-2000

RESULT 12

Qy   482 TCTATTCCGGGGACACCAATGCCCTGCCAGGCTCTGGTCCGGATGGGGAAA 531
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1985 TCTATTCCGGGGACACCAATGCCCTGCCAGGCTCTGGTCCGGATGGGGAAA 2034

Qy   422 TCGGGACACTGCAAGCACTCCGTTTGGCTCTGGTCGTCCACACCTCTGGAAAGTGG 481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1925 TCGGGACACTGCAAGCACTCCGTTTGGCTCTGGTCGTCCACACCTCTGGAAAGTGG 1984

Qy   362 AAAGATTGATCAGTTCGCTGCGTGTCCGAACATGTGAATTTCAGACCTGTCTGG 421
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1865 AAAGATTGATCAGTTCGCTGCGTGTCCGAACATGTGAATTTCAGACCTGTCTGG 1924

Qy   302 TCAGTATGATGATGATCTCGCCAAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTGTCCTTGCAGTGG 361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1805 TCAGTATGATGATGATCTCGCCAAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTGTCCTTGCAGTGG 1864

Qy   242 ACCAGCTCAAAGCCTGGCTCCACCAGTGCCAGGAGGTCCTCCACCACA 301
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1745 ACCAGCTCAAAGCCTGGCTCCACCAGTGCCAGGAGGTCCTCCACCACA 1804

Qy   182 GAGAACGCGCCCTTGGCATCTTTGGGAAAGCCCGTTCACCCTTTGCTGGTTGGTTGCCCCCA 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1685 GAGAACGCGCCCTTGGCATCTTTGGGAAAGCCCGTTCACCCTTTGCTGGTTGGTTGCCCCCA 1744

Qy   122 TGTTTGTGTCCCACCTGCACGCAGGATCACCACAGGGGCTTGCCAAGTATCTTGCTGCTCGAGA 181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1625 TGTTTGTGTCCCACCTGCACGCAGGATCACCACAGGGGCTTGCCAAGTATCTTGCTGCTCGAGA 1684

Qy   62 GGCAGCTGTGCCCTCATTACGGAGAGACCAGGTGACCAGGTTCCTGGGCACCCTGGCTGCTG 121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1565 GGCAGCTGTGCCCTCATTACGGAGAGACCAGGTGACCAGGTTCCTGGGCACCCTGGCTGCTG 1624

Qy   2 CACTTGTCAACAATAAGCCCCGACACGTCTCTGCTACTGGACTGGATGGTTGAGGGCACATTTG 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1505 CACTTGTCAACAATAAGCCCCGACACGTCTCTGCTACTGGACTGGATGGTTGAGGGCACATTTG 1564

ORIGIN
BASE COUNT   135 a   169 c   178 g   132 t

FEATURES   Location/Qualifiers
     source   1..761
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /clone="IMAGE:3945085"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
                     High quality sequence stop: 761.

                     DNA Sequencing by: Incyte Genomics, Inc.
                     Clone distribution: MGC clone distribution information can be
                     found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
                     Plate: LLCM802 row: p column: 14
                     Location/Qualifiers
                     High quality sequence stop: 761.

Qy   1479 GATGAAGATTCGAAAGTGCATGTCAGTTGCCACACTTGTCAACAATAAGCCCCGACACGTCTCTGCT 1538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GATGAAGATTCGAAAGTGCATGTCAGTTGCCACACTTGTCAACAATAAGCCCCGACACGTCTCTGCT 60

Qy   1539 ACTGGACTGGATGGTTGAGGGCACATTTGGGCAGGCTGTGCCCGTCATTACGGAGAGACCAGGTGGA 1598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACTGGACTGGATGGTTGAGGGCACATTTGGGCAGGCTGTGCCCGTCATTACGGAGAGACCAGGTGGA 120

Qy   1599 CAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGGATCACCACAC 1658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGGATCACCACAC 180

Qy   1659 GGGCTTGCCAAGTATCTTGCTGCTCGAGAGAGAAACGCGCCCTTGGCATCTTTGGGAAAGCCCGCT 1718
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGGCTTGCCAAGTATCTTGCTGCTCGAGAGAGAAACGCGCCCTTGGCATCTTTGGGAAAGCCCGCT 240

Qy   1719 TCACCCTTTGCTGGTTGGTTGCCCCCAACCAGCTCAAAGCCTGGCTCCACCAGTACCACAA 1778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TCACCCTTTGCTGGTTGGTTGCCCCCAACCAGCTCAAAGCCTGGCTCCACCAGTACCACAA 300

Qy   1779 CCAGTGCCAGGAGGTCCTGCACCACAACATCAGTATGATGATGATCAGTTCCTGCCAAATGCCTTCAGGAAGG 1838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCAGTGCCAGGAGGTCCTGCACCACAACATCAGTATGATGATGATCAGTTCCTGCCAAATGCCTTCAGGAAGG 360

Qy   1839 GGGCTGAGATCTCCAGTGTCCTTGCAGTGGAAAGATTGATCAGTTCGCTGCGTGTCCGAACATGTGA 1898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGGCTGAGATCTCCAGTGTCCTTGCAGTGGAAAGATTGATCAGTTCGCTGCGTGTCCGAACATGTGA 420

Qy   1899 TTTGGAAGAGTTCAGACCTGTCTGGGTGCGGCCACTGCAAGCATGCAGTTGCGTTGGCTGTGCGCT 1958
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 -TTTGGAAGAGTTCAGACCTGTCTGGGTGCGGCACTGCAAGCATGCAGTTGCGTTGGCTGTGCGCT 479

Qy   1959 GGTGCACACCTCTGGTCCGGAAAGTGGGTCTATTCCGGGGACACCAATGCCCTGCCAGGCTCT 2018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 GGTGCACACCTCTGGTCCGGAAAGTGGGTCTATTCCGGGGACACCAATGCCCTGCCAGGCTCT 539

Qy   2019 GGTCCGGATGGGGAAAGATGCCACACCTCCTGATACATGAAGCCACCCTGAAGATGGTTT 2078
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 GGTCCGGATGGGGAAAGATGCCACACCTCCTGATACATGAAGCCACCCTGAAGATGGTTT 599

Qy Best Local Similarity: 99.7%; Conservative: 0; Mismatches: 1; Indels: 1; Gaps: 1;
    Matches 698; Pred. No. 1.3e-267;
    Query Match 21.3%; Score 528; DB 140; Length 761;

BASE COUNT   172 a   206 c   219 g   164 t
ORIGIN

```
Qy    2079 ggaggaggaagcagtggaaagacacacagcacaagtcccaagccaatccagcgtgggat 2138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 GGAAGAGGAAGCAGTGGAAAAGACACACACCACCAACGTCCCAAGCCATCAGCGTGGGAT 659

Qy    2139 gcggatgaacgcggagttcattatgctgaaccacttcagc 2178
           ||||||||||||||||||||||||||||||||||||||||
Db     660 GCGGATGAACGCGGAGTTCATTATGTCTGAACCACTTCAGC 699

RESULT 13
BF196304
LOCUS        BF196304        526 bp    mRNA              EST        03-NOV-2000
DEFINITION   7n68a10.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3569586 3',
             similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION    BF196304
VERSION      BF196304.1   GI:11084107
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 526)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
             R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
             Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
             I.M.A.G.E. Consortium DNA Sequencing by: Washington University
             Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL, send email to:
             info@image.llnl.gov
             High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1..526
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3569586"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCGCGCGCGCACATTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      136 a    142 c    147 g    101 t
ORIGIN

        Query Match    21.2%;   Score 526;  DB 145;   Length 526;
        Best Local Similarity 100.0%;   Pred. No. 1.5e-266;
        Matches  526;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps   0;

Qy      90 gcgggccgccaagacccgctgcggcacctgcgcaacgcggaagagagaagacgcggaacc 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGGCCGCCGCCAAGGACCCCGTGCGGCACCTGCGCGAGAGAAGCGCGGAACCCGTCCGGG 60

Qy     150 gtgtctccggcgggcccaaacaccgttacctgcaggtggtggcagcgggttagccgggact 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGGCGGCGCCAAGGACCCCGCCAAAACCGTGTACTCGCAGGTGGTGGCGCACGGGGACTC 120

Qy     210 gggcggcgcgcctctacgtcttctccgagttcaaccggtatctcttcaactgtgtggagaagg 269
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     121 GGGCGCCGGCGCTCGGTCTTCTCCAGTTCAACACCGGTATCTCTTCAACTGTGTGGAGAAGG 180

Qy     270 cgttcagagacttcatgcaggaggacaagagttaaaggttgtctcgcctggacaacatattcct 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGTTCAGAGACTTCATGCAGGAGGACAAGTTAAAAGGTTGCTCGGCCTGGACAACATATTCCT 240

Qy     330 gacacgaaatgacactgttctcaatgttcggggggttaagtggaatgaatcttacttttaaagga 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GACACGAATGCACTGGTTCTCAATGTTCGGGGGGCTTAAGTGGAATGAATCTTACTTTTAAAGGA 300

Qy     390 aaccgggccttccaaagtgtgtactctttctgtgacctccacaactggaaaataacctcgaagc 449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AACCGGGCCTTCCAAAGTGTGTACTTTCTGTGACCTCCACAACTGGAAAAAAATACCTCGAAGC 360

Qy     450 aatcaaaaatatatttctggtccattgaaaggatagaactggctggtctgtgtgcggggcccacctctgc 509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AATCAAAAATATTTTCTGGTCCATTGAAAGGATAGAACTGGCTGGTCTGTGTGCGGGCCCACTCTGC 420

Qy     510 cccagaatacgaggatgaaacccatgacagtttaccagatcccccatacacacagtgaacagag 569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCCAGAATACGAGGATGAAACCCATGACAGTTTACCAGATCCCCCATACACACAGTGAACAGAG 480

Qy     570 gaggggaaagcaccaaccatggcagagtccagaaaaggccctccagc 615
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAGGGGAAAGCACCAACCATGGCAGAGTCCAGAAAAGGCCTCTCAGC 526

RESULT 14
AU138595
LOCUS        AU138595        696 bp    mRNA              EST        25-OCT-2000
DEFINITION   AU138595 PLACE1 Homo sapiens cDNA clone PLACE1008903 5', mRNA
             sequence.
ACCESSION    AU138595
VERSION      AU138595.1   GI:11000116
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 696)
AUTHORS      Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
             Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
             Isogai,T.
TITLE        HRI human cDNA project
JOURNAL      Unpublished (2000)
COMMENT      Contact: Takao Isogai
             Genomics Laboratory
             Helix Research Institute
             1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             Tel: 81-438-52-3951
             Fax: 81-438-52-3952
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
             Research Institute; cDNA library construction: Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..696
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1008903"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT      163 a    192 c    208 g    130 t     3 others
ORIGIN

        Query Match    21.2%;   Score 525;  DB 108;   Length 696;
        Best Local Similarity 100.0%;   Pred. No. 5.1e-266;
        Matches  525;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps   0;

Qy       1 atgtgggcgcttgctcgctcgcggtcgggccggacgcaccatgtcgcaggaggacgc 60
```

```
Db    73 ATGTGGGCGGGCTTTGTCCTGCTGCTGCTGGCGGTCCGGCCGCCGGACGGCACGTCCGCAGGGACGG 132
Qy    61 accatatacggcaggcaccgccgccgcggaaggaccccgcgctgggcactg              120
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 ACCATATCGGCAGGCACCG...                                            192
Qy   121 cgcacgcagaagaagcgcgggaccgtcggggtgctccggcgcggcccaaacaccgtgtactg    180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 GCCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGCCCAAACACCGTGTACCTG       252
Qy   181 caggtggtggcagcgggtagccggagctcggtgcgcgccgcctacgtcttctccgagttc     240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 CAGGTGGTGGCGGCACCGGGTAGCCGGGACTCGGGCGCCGCCTACGTCTTCTCCGAGTTC      312
Qy   241 aaccggtatctcttcaactgtggagaaggcgttcaggagacttcatgtcaggagcacaagtta  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGGAGACTTCATGTCAGGAGCACAAGTTA   372
Qy   301 aaggttgctcgcctggacaacatattcctggacgacgactggtctaaatgttggggggc     360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 AAGGTTGCTCGCCTGGACAACATATTCCTGGACGACGACTGGTCTAAATGTTGGGGGGC      432
Qy   361 ttaagtggaatgattcttacttaaaggaaaccgggcttccaaagtgtgtactctttga      420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCCGGGCTTCCAAAGTGTGTACTTTCTGA     492
Qy   421 cctccacaacctggaaaaaatacctcgaagcaatcaaaaatatttctggtccattgaaagga   480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 CCTCCACAACCTGGAAAAATACCTCGAAGCAATCAAAAATATTTCTGGTCCATTGAAAGGA    552
Qy   481 atagaactgggctgttgtcggccccccactctgccccagaaatacgaggat              525
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 ATAGAACTGGGCTGTTGCGGCCCCCCACTCTGCCCCAGAAATACGAGGAT               597
```

```
RESULT  15
BE794311    688    mRNA    EST    20-SEP-2000
LOCUS       BE794311   688 bp   mRNA   EST   20-SEP-2000
DEFINITION  601591442F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945538 5',
            mRNA sequence.
ACCESSION   BE794311
VERSION     BE794311.1  GI:10215496
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 688)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM804 row: c column: 11
            High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3945538"
                     /clone_lib="NIH_MGC_7"
                     /cell_line="MGC3"
                     /tissue_type="small cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    169 a  190 c  185 g  144 t
ORIGIN
```

```
Query Match      21.0%;  Score 521;  DB 140;  Length 688;
Best Local Similarity  99.8%;
Matches   64;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;
                       Pred. No. 6.6e-264;

Qy   770 tcaaagcaaaggaggatgggcctcccagttgggacagctgccatcgctcccatcattgtctg    829
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TCAAAGCAAAGGAGGATGGGCCTCCCAGTTGGGACAGCTGCCATCGGCTCCCATCATTGTCTG     60
Qy   830 ctgtcaaggacgggaaaaagcaatcactcatgaaggagaagaagaattttggctgaagctgt    889
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTGTCAAGGACGGGAAAAAGCCATCACTCATGAAGGAGAAGGAGATTTTGGCTGAAGCTGT    120
Qy   890 gtactcctccagatcctggtgtgctgctttcgttgtggtggtagaagtgtccaagctca      949
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTACTCCTCCAGATCCTGGTGCTGTGCTCTTTTGTTGTGGTGGTAGAAGTGTCCAAGCTCA    180
Qy   950 ttcaacccatctgtgagaatgccaaccttccagagggtaccaagagaaagggcaggtccccg   1009
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTCAACCCATCTGTGAGAAATGCCAACCTTCCAGAGGGTACCAAGAGAAAGGGCAGATGCCCCG    240
Qy  1010 tggcctttgtggttcacatggcccccacaatctctgtgcttgttggtggaccagcggtaccacgtgt  1069
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGGCCTTTGGTGGTTCACATGGCCCCCACAATCTGTGCTTGTTGGTGGACCAGCGGTACCAGT    300
Qy  1070 ggatggagagggttgggcctgacaccccagcacttggtcctgaatgagaactgtgcctcag   1129
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGATGGAGAGGGTTTGGGCCTGACACCCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAG    360
Qy  1130 ttcacaaccttcgcagcacaagattcaaacccagcctcaacctcatccaccggacatct   1189
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTCACAACCTTCGCAGCCACAAGATTCAAACCCAGCCTCAACCTCATCCACCGGACATCT    420
Qy  1190 tccccctcgctcacacagttccgtcgtgtaaggaagaaagaggggcccccaccccagttgtgccatg   1249
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TCCCCCTCGCTCACACAGTTTCCGCTGTAAGGAAGAGGGGCCCCA..CCTCAGTGTGCCCATGG    479
Qy  1250 ttcagggtgaatgcctcctcaagtacccagctccgtcccagggaggggggagtggcagaggggatg   1309
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 TTCAGGGTGAATGCCTCCTCAAGTACCCAGCTCCGTCCCAGGAGAGGGGGAGTGGCAGAGGGGATG    539
Qy  1310 ccattattcacttgcaatcctgaggaattcatagtgtgaggcgctgcagcttcccaacttcc   1369
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 CCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCC    599
Qy  1370 agcagaggacgtgcaggagtacagaggaggtgcgcaggacgggcc   1411
         |||||||||||||||||||||||||||||||||||||||||||||
Db   600 AGCAGAGGACGTGCAGGAGTACAGGAGGAGGTGCGCAGGACGCGCC    641
```

```
RESULT  16
BF797306    960    mRNA    EST    12-JAN-2001
LOCUS       BF797306   960 bp   mRNA   EST   12-JAN-2001
DEFINITION  602256926F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4340286 5',
            mRNA sequence.
ACCESSION   BF797306
VERSION     BF797306.1  GI:12102360
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
```

```
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9952 row: c column: 07
            High quality sequence stop: 705.
FEATURES             Location/Qualifiers
     source          1..960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4340286"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       230 a    286 c    276 g    168 t
ORIGIN

Query Match          20.9%;           Score 519;   DB 169;   Length 960;
Best Local Similarity 99.6%;           Pred. No. 7.7e-263;
Matches   669; Conservative   0; Mismatches   3;   Indels   0;   Gaps   0;

Qy  1362 caacttccagagagcgtgcaggagtacagagagggagtgcgcagaggacggccagcccagc 1421
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGAGTGCGCCAGAGGACGGCCCAGCCCAGC 60

Qy  1422 agagaaaaagaagtcagtacccagaaatcatcttccttggaacagggtctgccatcccgat 1481
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AGAGAAAAAGAAGTCAGTACCCAGAAATCATCTTCCTTGGAACAGGGTTCTGCCATCCCGAT 120

Qy  1482 gaagattcgaaatgtcagtgcacaacttgtcaacataagccccgaacacgtctctgctact 1541
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACT 180

Qy  1542 ggactgtggtgaggcaacattgggcagctgtgcgctcattacggagacacaggtggacag 1601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGACTGTGGTGAGGGCACGTTTTGGGCAGCTTGCGCGTCATTACGGAGACCACAGGTGGACAG 240

Qy  1602 ggtcctgggggaccctggctgctgtttgtgtcccacctgcacgcgagatacccacacggg 1661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGTCCTGGGCGACCCTGGCTCTGTGTTTGTGTCCCACCTGCACGCGAGATCACCACACGGG 300

Qy  1662 cttgccaagtatcttgctgcagagagagaacgcgccttggccatctttgggaaagccgcttca 1721
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTTGCCCAAGTATCTTGCTGCGAGAGAGAACCCGCGCCTTGGCCATCTTTGCGAAAGCCGCTTCA 360

Qy  1722 ccccttgctggttgttgccccccaaccagctccaaagccctggctccagcagtaccacacaacca 1781
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCCTTTTGCTCGGTTGCGTGTTGCCCCCACCAGCTCCAAAGCCTCCAGCAGTACCACACAACCA 420

Qy  1782 gtgtgcaaggaggtcctgccacacacatcagttatgaattccttcaggaagggc 1841
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTGTGCAAGGAGGTCCTGCACACAACATCAGTTATGATTCCTTCAGGAAGGGGC 480

Qy  1842 tgaatctccagtcctgcagtgaaagattgatcagttcgctgttcgcaacatgtgattt 1901
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGAGATCTCCAGTCCTGCACTGGAAAGATTGCATCAGTTCGCTGTTCGCAACATGTGATCT 540

Qy  1902 ggaggagtttcaagacctgtctggtgcggacaagcatgcgtttggctgtgtgcgctggt 1961
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GGAAGAGTTTCAGACCTGTGCTGTCGGCGCACTGCAAGCAATGCGATTGGCTGTGCGCTGGT 600
```

```
RESULT    17
LOCUS       AU143668
DEFINITION  AU143668 Y79AA1 Homo sapiens cDNA clone Y79AA1002297 5', mRNA
            sequence.
ACCESSION   AU143668
VERSION     AU143668.1    GI:11005189
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..670
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Y79AA1002297"
                     /clone_lib="Y79AA1"
                     /cell_type="retinoblastoma"
                     /cell_line="Y79"
                     /note="Vector: pME18SFL3"
BASE COUNT       157 a    190 c    193 g    127 t    3 others
ORIGIN

Query Match          20.4%;           Score 507;   DB 108;   Length 670;
Best Local Similarity 99.7%;           Pred. No. 1.7e-256;
Matches   607; Conservative   0; Mismatches   2;   Indels   0;   Gaps   0;

Qy     1 atgtgtgggccgcttgctcgtcgtcgttccgtccgggccggacgccgccaccatgtcgcagggagcgc 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    31 ATGTGGGCGCGTTTGCTCGCGTCGTGCGGCCTGCCGGACGGCCACCATGTCGGAGGGACGC 90

Qy    61 accatatcgcaggcaccccgcccggccccgagcgggccgcgcaaggaccgctcgggcaacctg 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    91 ACCATATCGCAGGCACCACCCGCCCGGCCCCGAGCGGGCCGCCAAGGACCCGCTCGGGCACCTG 150

Qy   121 cgcacgcgagagagaaggccggcggaaccgtcgggggtgactcgggggccccaaacacgtgtacctg 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 CGCACGCGAGAGAAGGCCGGCGGACCGTCGGGGGTGCTCGGGGCCCCAAACACCGTGTACCTG 180

Qy   181 caggtcgttgcagcgggtagcggggactcgggccgctctcacgtcttcctcgctgttc 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGGTCGTTGGCAGCGGGTAGCCGGGGACTCGGGCCGCCGCAAAACACCGGTCTTCCGAGTTC 240

Qy   241 aaccggtatctcttcaactgtggagaaggccgttcaggagactcatgcaggagagcacaagtta 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AACCGGTATCTCTTCAACTGTGGAGAAGCCGTTCAGGAGAGAGCTCATGCAGGAGAGCACAAGTTA 300
```

RESULT 18

```
Db  271 AACGGGTATTCTCTTCAACTGTGTGGAGAAGGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  301 aaggttgctcgctgtaccaacatattcctgacacgaatgcaactggtgtctaatgttgtggggc 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  331 AAGGTTGCTCGCCTGGCACAACATATTCCTGACACGAATGCAACTGGTGTCTAATGTTGTGGGGC 390
Qy  361 ttaagtggaatgattcttcctcacttaaaggaaaaccgggcttccaaagtgtgtgtactttctgga 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  391 TTAAGTGGAATGATTCTTCCTCACTTAAAGGAAAACCGGGCTTCCAAAGTGTGTGTACTTTCTGA 450
Qy  421 cctccacaaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  451 CCTCCACAAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGCCCATTGAAAAGGA 510
Qy  481 atagaactgctgtgcggccccaactctgcccccagaggaggatgaaaccatgacagtt 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  511 ATAGAACTGCTGTGCGGCCCCAACTCTGCCCCCAGAGGAGGATGAAACCATGACAGTT 570
Qy  541 taccagatccccatacacacagtgaacacagagagaggggggcacccaaaccatggcagaagtcca 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  571 TACCAGATCCCCATACACACAGTGAACACAGAGAGAGGGGGAAAGCACCACCAAACCATGGCAGAGTCCA 630
Qy  601 gaaaggcct 609
        |||||||||
Db  631 GAAAAGGCCT 639
```

LOCUS   BF057321   506 bp   mRNA   EST   16-OCT-2000
DEFINITION 7k19c01.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3475944 3'
            similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF057321
VERSION   BF057321.1   GI:10811217
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 506)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -400P from Gibco
            High quality sequence stop: 477.
FEATURES   Location/Qualifiers
     source   1..506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3475944"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAACTGGGAGGCGGCCGCGCGGCACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was

RESULT 19

BASE COUNT 133 a 133 c 139 g 101 t
ORIGIN   constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match 506;
Best Local Similarity 100.0%; Score 506; DB 143; Length 506;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy  108 gctgcggcacctgcgcacgcgagagagaaggcgcgaggaccgtcgggttgctcggcggcccaaa 167
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCTGCGGCACCTGCGCACGCGAGAGAGAAGGCGCGAGGACCGTCGGGTTGCTCGGCGGCCCAAA 60
Qy  168 caccgtgtacctgcaggtggtgtgggcagcgggtagccggagccgacggcatcctcatagt 227
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CACCGTGTACCTGCAGGTGGTGTGGGCAGCGGGTAGCCGGAGCCGACGGCGTGGGCGGCTCTACGT 120
Qy  228 cttctccgagttcaaaccggtatctcttcaaactgtggagaaggcggttcagaggactccatca 287
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTTCTCCGAGTTCAAACCGGTATCTCTTCAAACTGTGGGAGAAGGCGGTTCAGAGGACTCATGCA 180
Qy  288 gcagacaagttaaggttgctcgcctggacaacaatattccctgacacgaatgcaactggtc 347
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GGAGCACAAGTTAAAGGTTGCTCGCCTGGACAACAATATTCCCTGACACGAATGCAACTGGTC 240
Qy  348 taatgttggggcttaagtggaatgaattctttactttaaaggaaaaccgggcttccaaagtg 407
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TAATGTTGGGGCTTAAGTGGAATGAATTCTTTACTTTAAAGGAAAACCGGGCTTCCAAAGTG 300
Qy  408 tgtactttctggacctccacaactggaaaaatacctcgaagcaatcaaaatattttctgg 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGG 360
Qy  468 tccattgaaaggaataagaactagctgtgcgggccccaactctgccccagaatacgaggatga 527
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TCCATTGAAAGGAATAAGAACTAGCTGTGCGGGCCCCAACTCTGCCCCAGAATACGAGGATGA 420
Qy  528 aaccatgacagttaccagatcccatacacagtgaacacagaggaggaggggaaagcaccaac 587
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AACCATGACAGTTACCAGATCCCCATACACACAGTGAACACAGAGGAGGAGGGGAAAGCACCAAC 480
Qy  588 atggcagagagtccagaaaggcctctca 613
        ||||||||||||||||||||||||||||
Db  481 ATGGCAGAGAGTCCAGAAAAGGCCTCTCA 506
```

LOCUS   BF794577   953 bp   mRNA   EST   12-JAN-2001
DEFINITION 602255843F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4339187 5',
            mRNA sequence.
ACCESSION   BF794577
VERSION   BF794577.1   GI:12099631
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 953)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9949 row: e column: 12
            High quality sequence stop: 704.

FEATURES
    source   Location/Qualifiers
             1..953
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:4339187"
             /clone_lib="NIH_MGC_85"
             /tissue_type="lymphoma, cell line"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
             Average insert size 1.867 kb. Library enriched for
             full-length clones and constructed by Life Technologies.
             Note: this is a NIH_MGC Library."
BASE COUNT   214 a   255 c   271 g   213 t
ORIGIN

Query Match   20.0%;   Score 497;   DB 169;   Length 953;
Best Local Similarity   99.8%;   Pred. No. 3.3e-251;
Matches 547;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

```
Qy     599 cagaaaaggcctctcaagcagcaggctcagtccagagcgaatcttcaagactccgagtcgaatgaaa 658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12 cagaaaaggcctctcaagcagcaggctcagtccagagcgaatcttcaagactccgagtcgaatgaaa 71
Qy     659 atgagccacaccttccacatggtgttagccagagaagaggggtcaggagaactcctccctgg 718
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 atgagccacaccttccacatggtgttagccagagaagaggggtcaggagaactcctccctgg 131
Qy     719 tcgtagcttcattcgttagcttcacttaaagagaggaaacttcttggttgctcaaagcaa 778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 tcgtagcttcattcgttagcttcacttaaagagaggaaacttcttggttgctcaaagcaa 191
Qy     779 aggagagatgggcctccagtgggacagcgcatcgctccaccacattgctgctgtcaagg 838
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     192 aggagagatgggcctccagtgggacagcgcatcgctccaccatcattgctgctgtcaagg 251
Qy     839 acgggaaaagacatcactcatgaaggaagagagaattttgggctgaaaagcctgtgtactcctc 898
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     252 acgggaaaagacatcactcatgaaggaagagagaattttgggctgaaaagcctgtgtactcctc 311
Qy     899 cagatcctggtcgtctgctcttttgtgtggtagaaagcttcattcaaccca 958
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     312 cagatcctggtcgtctgctcttttgtgtggtagaaagcttcattcaaccca 371
Qy     959 tctgtgagaatgcaacccttcagagggtaccaagagaaagcagagatgccccccgtggccttgg 1018
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     372 tctgtgagaatgcaacccttcagagggtaccaagagaaagcagagatgccccccgtggccttgg 431
Qy    1019 tggttcaacatggccccagaatctgtcttgtgggacagcaagtgcaagtgggaga 1078
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 tggttcaacatggccccagaatctgtcttgtgggacagcaagtgcaagtgggaga 491
Qy    1079 ggtttgggcctgacaccacccagcacttggtcctgaatgtggcctcaagttcacaac 1138
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 ggtttgggcctgacaccacccagcacttggtcctgaatgtggcctcaagttcacaacc 551
Qy    1139 ttcgcaagc 1146
           |||||||||
Db     552 ttcgcagc 559
```

RESULT 20
BG519751
LOCUS   BG519751   884 bp   mRNA
DEFINITION   6025789002F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
           mRNA sequence.
ACCESSION   BG519751
VERSION   BG519751.1   GI:13515513
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
           High quality sequence stop: 859.

FEATURES
    source   Location/Qualifiers
             1..884
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:3503184"
             /clone_lib="NIH_MGC_19"
             /tissue_type="neuroblastoma"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
             EcoRI; cDNA made by oligo-dT priming. Directionally
             cloned into EcoRI/XhoI sites using the following 5'
             adaptor: GGCACGAG(G). Library constructed by Ling Hong
             in the laboratory of Gerald M. Rubin (University of
             California, Berkeley) using ZAP-cDNA synthesis kit
             (Stratagene) and Superscript II RT (Life Technologies).
             Note: this is a NIH_MGC Library."
             Plate: LLCM176 row: d column: 01
BASE COUNT   199 a   251 c   249 g   185 t
ORIGIN

Query Match   19.8%;   Score 491;   DB 154;   Length 884;
Best Local Similarity   100.0%;   Pred. No. 5e-248;
Matches 491;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    1315 attacttgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccacgcag 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 attacttgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccacgcag 332
Qy    1375 agcgtgcaggagttacaaggaggagggtgcgcaggagacggccagccaagagaagaaagaagt 1434
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 agcgtgcaggagttacaaggaggagggtgcgcaggagacggccagccaagagaagaaagaagt 392
Qy    1435 cagtaccaagaatcatcttcctgggaacaggtctgccatcccgatgaaagattcgaaat 1494
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     393 cagtaccaagaatcatcttcctgggaacaggtctgccatcccgatgaaagattcgaaat 452
Qy    1495 gtcagtgccacactgtcaacataagccccgacacgtctctgctactggtgttgag 1554
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     453 gtcagtgccacactgtcaacataagccccgacacgtctctgctactggactgtggtgag 512
Qy    1555 ggcacattgggcagcgtgccgtcaattacggagacaaggtggacaaggtcctgggcacc 1614
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     513 ggcacattgggcagcgtgccgtcaattacggagacaaggtggacaaggtcctgggcacc 572
Qy    1615 ctggctgctgtgtttgtcccaacctgcacgcagatcaccacacgggcttgccaagtatc 1674
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     573 ctggctgctgtgtttgtcccaacctgcacgcagatcaccacacgggcttgccaagtatc 632
Qy    1675 ttgctgcagagagaacgcgcctggcatcttggaaaagccgctcaccctttgctggtg 1734
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     633 ttgctgcagagagaacgcgcctggcatcttggaaaagccgctcaccctttgctggtg 692
Qy    1735 gttgcccccaaccagctcaaaggcctggctccagcgtaccacaacccagtgcaggaggtc 1794
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     693 gttgcccccaaccagctcaaaggcctggctccagcgtaccacaacccagtgcaggaggtc 752
Qy    1795 ctgcacccat 1805
           ||||||||||
Db    1795 ctgcacccat 1805
```

Db   753 CTGCACCACAT 763

RESULT 21
BG386348
LOCUS   BG386348   1012 bp   mRNA   EST   12-MAR-2001
DEFINITION 602455550F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583749 5',
            mRNA sequence.
ACCESSION   BG386348
VERSION   BG386348.1 GI:13279794
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1012)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1308 row: c column: 14
            High quality sequence stop:675.
FEATURES   Location/Qualifiers
     source   1..1012
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583749"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
                     /lab_host="DH10B (phage-resistant)"
BASE COUNT   246 a   273 c   281 g   210 t   2 others
ORIGIN

Query Match   19.3%;   Score 478; DB 153; Length 1012;
Best Local Similarity   99.8%;   Pred. No. 3.8e-241;
Matches 598; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy   896 ctccagatcctggtgctgctttgtggtgtagaatgtccagatgaaagcttcattcac 955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 CTCCAGATCCTGGTGCTGCTTTGTGGTGTAGAATGTCCAGATGAAAGCTTCATTCAAC 61

Qy   956 ccatctgtgagaatgccaccttcaagaggytaccacaggggaaggcagatgccccgtggcct 1015
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 CCATCTGTGAGAATGCCACCTTTCAAGAGGTACCACAAGGAAAGGCAGATGCCCCGTGGCCT 121

Qy   1016 tggtgttcacatggccccaacatctgtgcttgtgtgacagcaggtaccacagtggatgg 1075
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TGGTGGTTCACAAGGCCCCAGCANTCTGTGCTTGTGGACAGCAGGTACCACAGTGGATGG 181

Qy   1076 agaggtttgggcctgacaccag-cacttggttcctggatgctgcctcagttcac 1134
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   182 AGAGGTTTGGGCCTGACACCCAGTCACTTGGTTCCTGGATGCTGCCTCAGTTCAC 241

Qy   1135 aacctttcgcaagcacaacagattcaaacccagctcaaacctcatccacccggacactcttccc 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 AACCTTCGCAGCCACAGAATTCAAACCCAGCTCAAACCTCATCCACCCGGACACTCTTCCC 301

RESULT 22
AU138795
LOCUS   AU138795   657 bp   mRNA   EST   25-OCT-2000
DEFINITION AU138795 PLACE1 Homo sapiens cDNA clone PLACE1009311 5', mRNA
            sequence.
ACCESSION   AU138795
VERSION   AU138795.1 GI:11000316
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 657)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE   HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES   Location/Qualifiers
     source   1..657
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1009311"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT   158 a   176 c   192 g   125 t   6 others
ORIGIN

Query Match   19.2%;   Score 477; DB 108; Length 657;
Best Local Similarity 100.0%;   Pred. No. 1.2e-240;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 atgtgggcgcgtttgtctcgctcgcggtccgcgccggacgagcaccatgtcgcagggacgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 ATGTGGGCGCGTTTGTCTCGCTCGCGGTCCGCGCCGGACGAGCACCATGTCGCAGGGACGC 90

Qy   61 accatatcgcaggcaccccgccccgcggagggccgcgcaaggagaccccgctgcggcacctg 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 23 / RESULT 24 (rotated sequence alignment listing)

```
Qy  91  ACCAATATCGCAGGAGCCACCGCCGCGCGACCGGCCGAAGGACCACCGCGTGCGGGCACCTG  150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121  cgcacgcgagaggaagccggacgcgggcgggcgtcggggtgtctccggcgcccaaacaccgtgtacctg  180

Db 151  CGCACGGCAGAACGCCGGAGAGAACCGCGTGCGGGGCTCCGGCGGGCCCAAACACGGTGTACCTG  210
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 181  caggtggttgcgggaggtgggagcgccgcgcgcgcgtctacggtctctccgagttc  240

Db 211  CAGGTGGTTGGCACGCGGTAGCCGGGACTCGGGCGCCGCGCCTCAGCTCTTCTCCGAGTTC  270
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 241  aaccggtatctcttcaactgtgtggaagaggecgttcagagactccatgcagagcacaagtta  300

Db 271  AACCGGTATCTCTTCAACTGTGTGGAGAGACTCATCATGCAGAGCACAAGTTA  330
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 301  aaggttgctcgcctggacaacatattcctgacacgaagtgctctaatgttgggggc  360

Db 331  AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAAGTGCTCTAATGTTGGGGGC  390
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 361  ttaagtggaatgattcttacttttaaagaaggaaacccgggcttccaaaagtgtgttactttctggga  420

Db 391  TTAAGTGGAATGAATTCTTACTTTAAAGGAAAACCCGGGCGTTCCAAAGTGTGTACTTTCTGGA  450
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 421  cctccacaactggaaaaatacctcgaagcaatcaaaatatttttctggtccattgaaa  477

Db 451  CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTTCTGGTCCATTGAAA  507
```

LOCUS   BF311926   915 bp   mRNA   EST   21-NOV-2000
DEFINITION 601897767F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126864 5',
            mRNA sequence.
ACCESSION   BF311926
VERSION   BF311926.1 GI:11259697
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 915)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1017 row: f column: 17
            High quality sequence stop: 672.
FEATURES   Location/Qualifiers
     source   1..915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4126864"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note:this is a NIH_MGC library."
BASE COUNT   206 a   262 c   288 g   159 t
ORIGIN

RESULT 24

```
Query Match       19.1%;  Score 475; DB 147;  Length 915;
Best Local Similarity  99.8%;  Pred. No; 1.4e-239;
Matches  595; Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy 1074  ggagaggtttggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttca  1133
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2  GGAGAGGTTTGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCA  61

Qy 1134  caaccttcgcagccacaagattcaaacccagctcaaacctcaatccaaccgagacatcttccc  1193
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  62  CAACCTTCGGCAGCCACAAGAGATTCAAAACCCAGCTCAAACCTCAATCCACCGGGACATCTTCC  121

Qy 1194  cctgctcaaccagttccgctgtaaagaaggagggccccaccccaagtcgtgccaatggttca  1253
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 122  CCTGCTCAACCAGTTTCCGCTGTAAAGAAGGAGGGGCCCCACCCCAGTCGTGCCATGGTTCA  181

Qy 1254  gggtgaatgcctcctccaagtaccaagctccgtcccaggaggaggagtgaggagggatgtccat  1313
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 182  GGGTGAATGCCTCCTCCAAGTACCAAGCTCCGTCCCAGGAGGAGGAGTGGGAGGGGATGCCAT  241

Qy 1314  tattacttgcaatcctgaggaattcatagttgaggcggctgcagcttcccaaacttccagca  1373
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 242  TATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGGCTGCAGCTTCCCAAACTTCCAGCA  301

Qy 1374  gagcgtgcaggagtacaggagtgcgtcgcaggagacggcccagccacccagagaggaaagaag  1433
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 302  GACCGTGCAGGAGTACAGGAGGAGTGCGCAGGACGG--CCAGCCCCCAGCAGAGAAAAGAAG  360

Qy 1434  tcagtacccagaaaatcatcttccttggaacaggtgtctgccatcccgatgaagattcgaaa  1493
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361  TCAGTACCCAGAAAATCATCTTCCTTGGAACAGGTGTCTGCCATCCCGATGAAGATTCGAAA  420

Qy 1494  tgtcagttgccacacttgtcaaacataagccccgacacgtctctgctactgtggttga  1553
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421  TGTCAGTTGCCACACTTGTCAAACATAAGCCCCGACACGTCTCTGCTACTGTGGTTGA  480

Qy 1554  gggcacatttgggcagctgtgccgtcattacggagaccaggtggacaggggtcctgggcac  1613
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481  GGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCAC  540

Qy 1614  cctggcgctgctgtttggtcccaacctgcacgcagatcaccacacgggcttgcaa  1669
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541  CCTGGCGCTGCTGTTTGGTGTGTCCCAACCTGCACGCAGATCACCACACGGGCTTGCCAA  596
```

LOCUS   BE619259   823 bp   mRNA   EST   20-OCT-2000
DEFINITION 601473130F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5',
            mRNA sequence.
ACCESSION   BE619259
VERSION   BE619259.1 GI:9890197
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 823)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9636 row: k column: 08
            High quality sequence stop: 695.
FEATURES   Location/Qualifiers

COMMENT   Contact: Wilson RK
JOURNAL   Unpublished (1995)
TITLE   The WashU-Merck EST Project
             ,R., Williamson,A., Wohldmann,P. and Wilson,R.
             Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
             ,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J.,
             Hiller,L., Clark,L., Dubuque,T., Elliston,K., Hawkins,M., Holman
AUTHORS   1 (bases 1 to 531)
REFERENCE   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM   Homo sapiens
SOURCE   human.
KEYWORDS   EST.
VERSION   AA243700.1   GI:1874492
ACCESSION   AA243700
             SIS2-WTD1 INTERGENIC REGION. :, mRNA sequence.
             similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD PROTEIN IN
DEFINITION   zr68g08.s1 Soares_NhHPu_S1 Homo sapiens cDNA clone IMAGE:668606 3'
LOCUS   AA243700   531 bp   mRNA   EST   07-MAR-1997

RESULT   25

Matches 465; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
Best Local Similarity   100.0%; Pred. No. 2.8e-234;
Query Match   18.7%;   Score 465; DB 138; Length 823;

Qy 1628 ttgtgtccacctgcacgcagaatcacaccacacgggcttgccaagtatcttgctgcagagg 1687
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 TTGTGTCCCACCTGCACGCAGAATCACACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAG 134

Qy 1688 aacgcgccttggcatcttgggaaagccgacttcaccctttgctggtggttgcccccaacc 1747
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135 AACGCGCCTTGGCATCTTTGGGAAAGCCCGTTCACCCTTTGCTGGTGGTTGCCCCAACC 194

Qy 1748 agctcaaagcctggctccagcagtaccacacaacaagtgccaggagggtcctgcaacatca 1807
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195 AGCTCAAAGCCTGGCTCCAGCAGTACCACAACAAGTGCCAGGAGGGTCCTGCAACATCA 254

Qy 1808 gtatgatcctgccaaatgccctcaggagaggagtctccagtcctgcagtgagg gaa 1867
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255 GTATGAATTCCTGCCAAATGCCTTCAGGAGAGGGGGCTGAGAGATCTCCAGTCCTGCAGTGGGAA 314

Qy 1868 gattgatcagttcgctgttgcgaacatgtgattggaagagtttcagcacctgtctgtgtgc 1927
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   315 GATTGATCAGTTCGCTGTTGCGAACATGTGATTGGAAGACATTTCAGCACCTGTCTGTGGTGC 374

Qy 1928 ggcaactgcaagcatgcgtttggctgttgcgctggtgtgcacacctctggctggaaagtggtct 1987
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   375 GGCAACTGCAAGCATGCGTTTGGCTGTGTGCGCTGGTGTGCACACCTCTGGCTGGAAAGTGGTCT 434

Qy 1988 attccgggacacacatgccctgcgagggctctgtgtcccggatgggaaagatgccacccctcc 2047
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   435 ATTCCGGGGACACACCATGCCCTGCGAGGGTCTGGTCCCGGATGGGGAAAAGATGCCACCCTCC 494

Qy 2048 tgatacatgaagccaccctggaagagatggttgttggaaagagggaagcag 2092
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   495 TGATACAATGAAGCCACCCTGGAAGAGATGGTTTTGGAAGAGAGGAAAGCAG 539

FEATURES   Location/Qualifiers
source   1..531
             /organism="Homo sapiens"
             /db_xref="GDB:556573"
             /db_xref="taxon:9606"
             /clone="IMAGE:668606"
             /clone_lib="Soares_NhHMPu_S1"
             /tissue_type="pooled human melanocyte, fetal heart, and
             pregnant uterus"
             /lab_host="DH10B"
             /note="Organ: mixed (see below); Vector: pT7T3D-Pac
             (Pharmacia) with a modified polylinker; Site_1: Not I;
             Site_2: Eco RI; Equal amounts of plasmid DNA from three
             normalized libraries (melanocyte 2NbHM, pregnant uterus
             NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
             were made in vitro. Following HAP purification, this DNA
             was used as tracer in a subtractive hybridization
             reaction. The driver was PCR-amplified cDNAs from pools of
             5,000 clones made from the same 3 libraries. The pools
             consisted of I.M.A.G.E. clones 260232-265223,
             340488-345479, and 484488-489479."
             High quality sequence stop: 466.
BASE COUNT   144 a   137 c   143 g   107 t
ORIGIN

Matches 458; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
Best Local Similarity   100.0%; Pred. No. 1.4e-230;
Query Match   18.5%;   Score 458; DB 4; Length 531;

Qy   201 ccgggactcgggcgccgcgctctacgtctctccgcagttcaaccgtgtatctcttcaactg 260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 CCGGGACTCGGGCGCCGCGTCTACGTTCTTCTCCCAGTTCAAACCCGTATCTTCTTCAAACTG 133

Qy   261 tggagaaggcgttcaggagactcatgcaggagacacaagttaaaggttgtcgcctggacaa 320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134 TGGAGAAGGCGTTCAGAGACTCATGCAGGAGACACAAGTTAAAGGTTGCTGCCTGGACAA 193

Qy   321 catattccctggacacgaatgcactggttcttaaatgttgtggggggttaagtggaatgatccttac 380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 CATATTCCCTGGACACGAATGCCACTGGTTCTTAAATGTTGGGGGGGGCTTAAGTGGAATGATTCTTAC 253

Qy   381 tttaaagggaaacccgggcttccaaagtgtgtactttcctggacctcacacaactggaaaata 440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254 TTTAAAGGGAAAACCGGGCCTTCCAAAGTGTGTTACTTTCCTGGACCTCCACAACTGGAAAAATA 313

Qy   441 cctcgaagcaatcaaaatatttttcggtccattgaaagaaagaataagaacctggctgtgcggcc 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 CCTCGAAGCAATCAAAATATTTTTCTGGTCCATTGAAAGAAGAATAGAACTGGCTGTGTCGGCC 373

Qy   501 ccactctgccccagaatacgagggatgaaaccatgacagttaccagagatcccatacacag 560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 CCACTCTGCCCCAGAATACGGAGGATGAAAACCATGACAGTTACCAGATCCCATACACAG 433

Qy   561 tgaacaagagagggaagcacccaaaccatggcaagagtccagaaaggcctctcagcaggct 620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 TGAACAGAGAGGGAAAGCACCCAAACCATGGCAGAGTCCAGAAAGGCCCTCTCAGCAGGCT 493

Qy   621 cagtccagaggatcttcagacttccgagtcgaatgaaa 658
         ||||||||||||||||||||||||||||||||||||||
Db   494 CAGTCCAGAGCGATCTTCAGACTCCGAGTCGAATCAAA 531

---

source   1..823
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:3876223"
             /clone_lib="NIH_MGC_68"
             /tissue_type="large cell carcinoma"
             /lab_host="DH10B (phage-resistant)"
             /note="Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
             Seq primer: -41m13 fwd. ET from Amersham
             Average insert size 1.8 kb. Library constructed by Life
             Technologies."
BASE COUNT   185 a   219 c   259 g   160 t
ORIGIN

FEATURES
source

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -41m13 fwd. ET from Amersham

```
RESULT 26

BG471513
LOCUS       BG471513        611 bp    mRNA              EST      21-MAR-2001
DEFINITION  602513052F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4635685 5',
            mRNA sequence.
ACCESSION   BG471513
VERSION     BG471513.1  GI:13403788
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1393 row: g column: 14
            High quality sequence stop: 590.
FEATURES             Location/Qualifiers
     source          1..611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4635685"
                     /clone_lib="NIH_MGC_16"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5,
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC library."
BASE COUNT      145 a    171 c    161 g    134 t
ORIGIN

Query Match          18.3%;    Score 453;    DB 154;    Length 611;
Best Local Similarity   99.6%;    Pred. No. 6.1e-228;
Matches 553;    Conservative    0;    Mismatches    2;    Indels    0;    Gaps    0;

Qy     705 ggactcttcctcggtcgtagcttcatctgtaaggttcacttaaagagagagaaacttctt 764
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19 ggactcttcctcggtcgtagcttcatctgtaaggttcacttaaagagagagaaacttctt 78

Qy     765 ggtgctcaaagcaaagagaatgggcctcccagcttgggacagctgccatcgctccatcat 824
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 ggtgctcaaagcaaagagaatgggcctcccagcttgggacagctgccatcgctccatcat 138

Qy     825 tgctgctgtcaaggacggggaaaagcatcactcatgaaggaagagagagatttggctgaaga 884
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 tgctgctgtcaaggacggggaaaagcatcactcatgaaggaagagagagatttggctgaaga 198

Qy     885 gctgtgtactcctccagatcctggtggctgcttttgtgtggtggtagaagtgtccagatgaaag 944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 gctgtgtactcctccagatcctggtggctgcttttgtgtggtggtagaagtgtccagatgaaag 258

Qy     945 cttcaattcaaaccatctgtgagaatgcacctttcaagaggtaccaaaggaaaggcagagtgc 1004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 cttcaattcaaaccatctgtgagaatgcacctttcaagaggtaccaaaggaaaggcagagtgc 318

Qy    1005 ccccgtggccttggttgttcacatggcccccagcatctgtctgttggtgacagcaggtacca 1064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ccccgtggccttggttgttcacatggcccccagcatctgtctgttggtgacagcaggtacca 378

Qy    1065 gcagtggatggagaggtttgggcctgacacccagcacttggtcctgaatgagaaactgtgc 1124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     379 gcagtggatggagaggtttgggcctgacacccagcacttggtcctgaatgagaaactgtgc 438

Qy    1125 ctcagttcacaaccttcgcagccacaagattcaaaccccagcagctcaaccttcatccacccgga 1184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ctcagttcacaaccttcgcagccacaagattcaaaccccagcagctcaaccttcatccacccgga 498

Qy    1185 catctcccctgctcaccagttccgctgtaagaggaggaggagggccccacccccagtgtgcc 1244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 catctcccctgctcaccagttccgctgtaagaggaggaggagggccccacccccagtgtgcc 558

Qy    1245 catggttcaaggtga 1259
           |||||||||||||||
Db     559 catggttcaaggtga 573


RESULT 27

AI200296
LOCUS       AI200296        451 bp    mRNA              EST      14-OCT-1998
DEFINITION  qf86b12.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1756895 3' similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8
            KD PROTEIN IN SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AI200296
VERSION     AI200296.1  GI:3752902
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 442.
FEATURES             Location/Qualifiers
     source          1..451
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1756895"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGACGGACGCGCGCAATTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT      110 a    120 c    128 g     93 t
ORIGIN

Query Match          18.2%;    Score 451;    DB 17;    Length 451;
Best Local Similarity  100.0%;    Pred. NO. 6.9e-227;
Matches 451;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy      90 gcggccggccgcaaggaccccgctgcggccgccgaggaagcggaggaacgcggaccgtcggg 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gcggccggccgcaaggaccccgctgcggccgccgaggaagcggaggaacgcggaccgtcggg 60
```

RESULT 28

BE867512

LOCUS   BE867512   812 bp   mRNA   EST   20-OCT-2000
DEFINITION 601143010F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847226 5',
            mRNA sequence.
ACCESSION   BE867512
VERSION   BE867512.1   GI:10316288
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 812)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9561 row: c column: 03
            High quality sequence stop: 686.
FEATURES   Location/Qualifiers
     source   1..812
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3847226"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."
BASE COUNT   200 a   216 c   220 g   176 t
ORIGIN

RESULT 29

BF433563

LOCUS   BF433563   498 bp   mRNA   EST   30-MAR-2001
DEFINITION 7g55b07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3702108 3'
            similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF433563
VERSION   BF433563.1   GI:11445735
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 498)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -400P from Gibco

FEATURES   Location/Qualifiers
  source   1..498
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3702108"
                  /clone_lib="NCI_CGAP_Lu24"
                  /tissue_type="carcinoid"
                  /lab_host="DH10B"
                  /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                  modified polylinker; Plasmid DNA from the normalized
                  library NCI_CGAP_Lu5 was prepared, and ss circles were
                  made in vitro. Following HAP purification, this DNA was
                  used as tracer in a subtractive hybridization reaction.
                  The driver was PCR-amplified cDNAs from a pool of 5,000
                  clones made from the same library (cloneids
                  1414920-1417991 and 1520904-1522439). Subtraction by Bento
                  Soares and M. Fatima Bonaldo."
BASE COUNT   132 a   130 c   131 g   105 t
ORIGIN

Query Match   18.0%; Score 447; DB 148; Length 498;
Best Local Similarity 99.8%; Pred. No. 9.1e-225;
Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   142 ccgtcgtgggtgctccggcgcggcccaaaacacgtgtacctgcaggtggtggaggcgggtagc 201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CCGTCGGGGTGCTCCGGCGCGGGCCCAAACACCGTGTACCTGCAGGTGGTGGAGGCGGGTATC 60

Qy   202 cgggactcgggcgcgcgcgctctacgtctctcccgagttcaaccgtatctctcaactgt 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CGGGACTCGGGCGCGCGCGCTCTACGTCTTCTCCGAGTTCAACCGTATCTCTTCAACTGT 120

Qy   262 ggagaaggcgttcagagagactcatgcaggagcacaagttaaaggttgctcgcctggacaac 321
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGAGAAGGCGGTTCAGAGACTCATGCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAAC 180

Qy   322 atattcctgacacggatgcactggttctaatgttggggcttaagtgaatgagttcttact 381
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATATTCCTGACACGGATGCACTGGTTCTAATGTTGGGGCTTAAGTGAATGAGTTCTTACT 240

Qy   382 ttaaaggaaaccgggccttccaaaagtgtgtactttctggacctccacaactggaaaaatac 441
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TTAAAGGAAACCGGGCCTTCCAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATAC 300

Qy   442 ctcgaagcaatcaaaaatatttctggtccattgaaaggaataagaacttggctgtgtggccc 501
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTCGAAGCAATCAAAAATATTTCTGGTCCATTGAAAGGAATAAGAACTGGCTGTGTCGGCCC 360

Qy   502 cacttgccccagaataccgaggatgaacaatgacagttacccagatcccatacacagt 561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CACTCTGCCCCAGAATACGAGGATGAACAATGACAGTTACCCAGATCCCATACACAGT 420

Qy   562 gaacagagaagggaggggaaagcaccacaccatggcagagtccagaaggcctccacaggctc 621
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAACAGAGAGGAGGGGAAAGCACCACACCATGGCAGAGTCCAGAAAAGGCCCTCAGCAGGCTC 480

Qy   622 agtccagaggcgatctcca 639
          |||||||||||||||||||
Db   481 AGTCCAGAGACGGATCTTCA 498

RESULT 30
LOCUS   BE260626   938 bp   mRNA   EST   26-OCT-2000
DEFINITION   601146116F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161691 5',
                mRNA sequence.
ACCESSION   BE260626
VERSION   BE260626.1 GI:9132065
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 938)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM119 row: o column: 04
             Note: this is a NIH_MGC Library."
             High quality sequence stop: 621.
FEATURES   Location/Qualifiers
  source   1..938
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3161691"
                  /clone_lib="NIH_MGC_19"
                  /tissue_type="neuroblastoma"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                  EcoRI; cDNA made by oligo-dT priming; Directionally
                  cloned into EcoRI/XhoI sites using the following 5'
                  adaptor: GGCACGAG(G). Library constructed by Ling Hong
                  in the laboratory of Gerald M. Rubin (University of
                  California, Berkeley) using ZAP-cDNA synthesis kit
                  (Stratagene) and Superscript II RT (Life Technologies).
BASE COUNT   190 a   263 c   283 g   202 t
ORIGIN

Query Match   18.0%; Score 447; DB 165; Length 938;
Best Local Similarity 100.0%; Pred. No. 9.4e-225;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   916 tttgtggttagaatgtccagatgaaagcttcattcaaccatctgtgagaatgccacc 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135 TTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGAGAATGCCACC 194

Qy   976 tttcagagggtaccaaggagaaagcagagagtgccccccgtggccttggtggttcacatggccca 1035
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195 TTTCAGAGGTACCAAGGAAAAGCCAGAATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCA 254

Qy   1036 gcatctgtgcttgtgacacagcaggtaccacagcagtgaatgagagagagtttgggcctgacac 1095
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255 GCATCTGTGCTTGTGGACACAGCAGGTACCACAGCAGTGGATGGAGGGTTTGGGCCTGACAC 314

Qy   1096 cagcaacttggtcctgaatgagaactgtgccccagttcacaacaccttcgcagccacaagatt 1155
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   315 CAGCACTTGGTCCTGAATGAGAACTGTGCCCCAGTTCACAACCTTCGCAGCCACAAGATT 374

Qy   1156 caaaccaagctcaaacctcaccacccggacacatcttcccccctgctcaccagttcccgctgt 1215
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   375 CAAACCCAAGCTCAAACCTCATCCACCCGGACACATCTTCCCCCTGCTCACCAGTTCCCGCTGT 434

Qy   1216 aagaaggagggcccccaccccctccagtgtgcccatggttcaggtgaatgcctctcaagtac 1275
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   435 AAGAAGGAGGGCCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAGTAC 494

Qy   1276 cagctccgtccagaagggagggagtggcagaagggatgccattattacttgcaatcctgagaa 1335
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   495 CAGCTCCGTCCAGGAGGGAGTGGCAGAGGATGCCATTATTACTTGCAATCCTGAGGA 554

Qy   1336 ttcatagttgaggcggctgcaagcttccc 1362
          |||||||||||||||||||||||||||||
Db   555 TTCATAGTTGAGGCGGCTGCAGCTTCCC 581

RESULT 31

LOCUS	AW575677 493 bp mRNA EST 15-MAR-2000
DEFINITION	UI-HF-BM0-adl-b-07-0-UI.s1 NIH_MGC_38 Homo sapiens cDNA clone
	IMAGE:3061957 3', mRNA sequence.
ACCESSION	AW575677
VERSION	AW575677.1 GI:7247216
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 493)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.
	Email: cgapbs-r@mail.nih.gov
	Oligo-dT track not found, but I site shown in beginning of sequence
	is likely internal to the message. Tissue Procurement: Louis M.
	Staudt, M.D., Ph.D.
	cDNA Library Preparation: M.B. Soares Lab
	cDNA Library Arrayed by: M.B. Soares Lab
	DNA Sequencing by: M.B. Soares Lab
	Clone distribution: MGC clone distribution information can be
	found through the I.M.A.G.E. Consortium/LLNL at:
	www-bio.llnl.gov/bbrp/image/image.html
	Seq primer: M13 Forward
	POLYA=No.
FEATURES	Location/Qualifiers
source	1..493
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3061957"
	/clone_lib="NIH_MGC_38"
	/tissue_type="lymph"
	/cell_type="germinal center B cells"
	/cell_line="MGC85"
	/lab_host="DH10B (LTI)"
	/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
	Constructed from size fractionated cytoplasmic mRNA
	(2.5-3.5kb). Directionally cloned. Cells provided by Louis
	M. Staudt, Ph.D. Library preparation by Maria de Fatima
	Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT	126 a 131 c 138 g 97 t 1 others
ORIGIN	

```
Query Match         17.8%;  Score 442;  DB 118;  Length 493;
Best Local Similarity   99.8%;
Matches  492;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      91 cggcccgcgcaaggaccgctgggcacctgcgcacgcgcagaggaagcgcggaacgtcgggg 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGGCCCGCGCAAGGACCCGCTGGGCACCTGCGCACGCGCAGAGGAAGCGCGGAACGTCGGGG 60

Qy     151 tgctccgggcggcccaaacaccgtgtacctgcagtgtgttgcagcgtggtagccgggactcg 210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCTCCGGGCGGGCCCAAACACCGTGTACCTGCACGTGGTGGCCACCGGGTGAGCCGGGACTCG 120

Qy     211 ggcgcgcgcgcttacgtctttcccgagttcaaccgtgtatctctttcaactgtggagaagc 270
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGCGCGCGCGCTTACGTCTTTCCCGAGTTCAACCGGTATCTCTTCTTCAACTGTGGAGAAGGC 180

Qy     271 gttcagagaactcatgcagagcacaagttaaaggttcgctccgtggacaacatattcctg 330
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTTCAGAGACTCATGCAGAGGACAAGTTAAAGGTTCGCTCCGTGGACAACATATTCCTG 240

Qy     331 acacgaatgcactggtctaatgtgggggcttaagtgaatgacttcttactttaaggaa 390
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACACGAATGCACTGGTCTAATGTGTGGGGCTTAAGTGGAATGAGTTCTTACTTTAAAGGAA 300

Qy     391 accgggcttccaaagtgtgtacttctcggacctccacaactggaaaaatacctcgaagca 450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```
```
Db     301 ACCGGGCGCTTCCAAAGTGTGTACTTTCTGGACCTCCACAACTGTGGAAAAAAATACCTCGAAGCA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     451 atcaaaaatattttctggtccattgaaaggaatagaaaactggcctgtcgcggcccacctctgcc 510
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATCAAAAATATATTTTCTGGTCCATTGAAAGGGAATAGAAAACTGGCGTGTCGCGGCCCACTCTGCC 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     511 ccagaaatacgaggtgatgaaacctgacagttacaggatcccccatacacagtgaacagagg 570
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCAGAAATACGAGGAGATGAAAACCATGACAGTTTACCAGATCCCCATACACAGTGAACAGAGG 480
           |||||||||||||||||||||||||||
Qy     571 agggagaagcacc 583
           ||||||||||||||
Db     481 AGGGGAAAAGCACC 493
```

RESULT 32

LOCUS	BE304720 612 bp mRNA EST 13-JUL-2000
DEFINITION	601106236F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:3349304 5',
	mRNA sequence.
ACCESSION	BE304720
VERSION	BE304720.1 GI:9176150
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 612)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.
	Email: cgapbs-r@mail.nih.gov
	Tissue Procurement: ATCC
	cDNA Library Preparation: Ling Hong/Rubin Laboratory
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: NIH Intramural Sequencing Center
	Clone distribution: MGC clone distribution information can be
	found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
	Plate: LLCM143 row: h column: 09
	High quality sequence start: 21
	High quality sequence stop: 609.
FEATURES	Location/Qualifiers
source	1..612
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3349304"
	/clone_lib="NIH_MGC_15"
	/tissue_type="adenocarcinoma cell line"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
	EcoRI; cDNA made by oligo-dT priming. Directionally
	cloned into EcoRI/XhoI sites using the following 5'
	adaptor: GGCACGAG(G). Size-selected >500bp for average
	insert size 1.8kb. Library constructed by Ling Hong in
	the laboratory of Gerald M. Rubin (University of
	California, Berkeley) using ZAP-cDNA synthesis kit
	(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT	143 a 178 c 161 g 130 t
ORIGIN	

```
Query Match         17.7%;  Score 438;  DB 166;  Length 612;
Best Local Similarity  100.0%;
Matches  438;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     991 ggaaaggcagatgccccctggccttggttgttcacatggcccacagcatctgcttgtg 1050
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     108 GGAAAGGCAGATGCCCCCTGGCCTTGGTTGTTCACATGGCCCACAGCATCTGCTTGTG 167

Qy    1051 gacagcaggtaccagcagtggatggagaggtttgggcctgacacaccagcacttggtcctg 1110
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   168 GACACCAGTGTACCACCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCACTTGTCCTG 227
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1111 aatgagaactgtgcctcagttcacacacctcgcagccacaagattcaaaaccagctcaac 1170

Db   228 AATGAGAACTGTGCCTCAGTCACAACCTTCGCACCACAAGATTCAAACCCAGCTCAAC 287
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1171 ctcatccaccccggacatcttccccccctgctccgcctgtaagagaaggaggggccc 1230

Db   288 CTCATCCACCCGGACATCTTCCCCCTGCTCCAGTTCCGCTGTAAGAGGAGGGCCCC 347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1231 accctcagtgtgccatggttcaagggtgaatgcctcctccagctaccagctccgtcccagg 1290

Db   348 ACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGG 407
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1291 agggagtggcagaggaggaggatgccattattactttgcatctgagaagttcatagttgaagcg 1350

Db   408 AGGGAGTGGCAGAGGAGGATGCCCATTATTACTTGCAATCCTGAGAAGTTCATAGTTGAGGCG 467
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1351 ctgcagcttccccacacttccagcacgacagagacgtgcagaagtacaggagagtgcgcaggacggc 1410

Db   468 CTGCAGCTTCCCACACTTCCAGCACGACAGAGGACGTGCAGGAGTACAGGAGAGTGCGCAGGACGGC 527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1411 ccagccccagcagaaggaa 1428

Db   528 CCAGCCCCAGCAGAGAAA 545
             ||||||||||||||||
```

RESULT 33
AA455121

LOCUS   AA455121   493 bp   mRNA   EST   06-JUN-1997
DEFINITION zx78c04.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
            IMAGE:809862 3', similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8
            KD PROTEIN IN SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AA455121
VERSION   AA455121.1   GI:2177897
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 493)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Mara,M., Martin,J., Moore,B.,
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            , T., Waterston,R. and Wilson,R.R.
TITLE   WashU-Merck EST Project 1997
JOURNAL   Unpublished (1997)
COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Seq primer: -41m13 fwd ET from Amersham
            High quality sequence stop: 447.
FEATURES   Location/Qualifiers
     source   1..493
                     /organism="Homo sapiens"
                     /db_xref="GDB:6039680"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:809862"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCGGCCGCCGGTTTTTTTTTTTTTTTTTTTT 3'],
```

BASE COUNT   131 a   128 c   130 g   104 t
ORIGIN

```
Query Match            17.6%;   Score 436;  DB 7;   Length 493;
Best Local Similarity  100.0%;  Pred. No. 6e-219;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy   202 cgggacctcgggcgccggccgctctacgtctttcctccgagttccaccggtatctcttcaactgt 261
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    58 CGGGACTCGGGCGCCGGCCGCTCTACGTCTTTCCTCCGAGTTCCACCGGTATCTCTTCAACTGT 117

Qy   262 ggagagaaggcgttcaggagactcatgcaggagcacaagttaaagttgctgctcctggacaac 321
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118 GGAGAAGGCGGTTCAGGAGACTCATGCAGGAGCACAAGTTAAAGGTTGCTGCTGGACAAC 177

Qy   322 atattcctgacacgaatgcactggttctcaatgttgggggcttaagtggaatgattcttact 381
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178 ATATTCCTGACACGAATGCACTGGTTCTCAATGTTGGGGGCTTAAGTGGAATGATTCTTACT 237

Qy   382 ttaaagggaaaaccgggcttccaaagtgtgtacttctgggaccctccacaactggaaaaatac 441
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   238 TTAAAGGAAAAACCGGGCTTCCAAAGTGTGTACTTTCTGGGACCCTCCACAACTGGAAAAATAC 297

Qy   442 ctcgaagcaatcacaaaatattttcctggtccaatgaaaggaaaatagaactgggctgtcgcgccc 501
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   298 CTCGAAGCAATCACAAAATATTTTCTGGGTCCAATGAAAGGAAAATAGAACTGGGCTGTCGCGCCC 357

Qy   502 cactctgcccagaaatacgaggagatgaaaacctgacagttacccagatcccacatacacagt 561
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   358 CACTCTGCCCAGAAATACGAGGAGATGAAAACCATGACAGTTACCCAGATCCCCATACACAGT 417

Qy   562 gaacacagagaggggaaagcaccaacctgtgcagagtccagaaaggcctctccagcaggctc 621
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   418 GAACACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCAGAAAGGCCTCTCAGCAGGCTC 477

Qy   622 agtccagacggatctt 637
             ||||||||||||||||
Db   478 AGTCCAGACGGATCTT 493
```

RESULT 34
AI468143

LOCUS   AI468143   421 bp   mRNA   EST   30-MAR-1999
DEFINITION tf92g05.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2106776 3'
            similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD PROTEIN IN
            SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AI468143
VERSION   AI468143.1   GI:4330233
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 421)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project
            (CGAP), Tumor Gene Index
JOURNAL   Unpublished (1998)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
```

## Left column

```
FEATURES
     source
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert Length: 733   Std Error: 0.00
              Seq primer: -40UP from Gibco.
              Location/Qualifiers
              1..421
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:2106776"
              /clone_lib="NCI_CGAP_Brn23"
              /tissue_type="glioblastoma (pooled)"
              /lab_host="DH10B"
              /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
              modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
              strand cDNA was primed with a Not I - oligo(dT) primer [5'
              TGTTACCAATCTGAAGTGGGACGGCGGCCGCATANCTTTTTTTTTTTTTTTTTTTTT
              T 3']; double-stranded cDNA was ligated to Eco RI
              adaptors (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of the modified pT7T3 vector.
              Library is normalized, and was constructed by Bento
              Soares and M.Fatima Bonaldo."
BASE COUNT         114 c      121 g       88 t
ORIGIN       98 a
```

```
RESULT 35
LOCUS       AW206103      536 bp    mRNA           EST   02-DEC-1999
DEFINITION  UI-H-BI1-afx-f-12-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
            IMAGE:2723471 3', mRNA sequence.
ACCESSION   AW206103
VERSION     AW206103.1   GI:6505577
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
```

```
Query Match        17.0%;   Score 421;  DB 20;  Length 421;
Best Local Similarity  100.0%;  Pred. No. 5.1e-211;
Matches   421;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     90 gggcgccgcaagagaccgctgcgcacgcgaggagaagcgcggagacgtcggg 149
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gcggcgccgcaagagaccgctgcgcacgcgaggagaagcgcggagacgtcggg 60

Qy    150 gtgctccggcgggcccaaacaccgtgtacctgcaggtgtggcagagccggactc 209
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 gtgctccggcgggcccaaacaccgtgtacctgcaggtgtggcagagccggactc 120

Qy    210 gggcgccgcgctctacgtctcttccgagttcaaccgtatctcaactgtggagaag 269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gggcgccgcgctctacgtctcttccgagttcaaccgtatctcaactgtggagaag 180

Qy    270 cgttcagagactcatgcaggaggacaagttaaaggttgctcgcctggacaacatattcct 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cgttcagagactcatgcaggaggacaagttaaaggttgctcgcctggacaacatattcct 240

Qy    330 gacacgaatgcactggtctaaatgttggggcttacagtggaatgatgatcttactttaaagga 389
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gacacgaatgcactggtctaaatgttggggcttacagtggaatgatgatcttactttaaagga 300

Qy    390 aaccgggcttcaaagtgtgtactttctggacctccacaactggaaaaatacctcgaagc 449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 aaccgggccttcaaagtgtgtactttctggacctccacaactggaaaaatacctcgaagc 360

Qy    450 aatcaaaatattttctgtccattgaaagagaataagaactggctgtgtgcgggccccactctgc 509
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AATCAAAATATTTTCTGTCCATTGAAAGGAATAAGAACTGGCTGTGTGCGGGCCCCACTCTGC 420

Qy    510 c 510
          |
Db    421 c 421
```

Right column

```
REFERENCE   1 (bases 1 to 536)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI-CGAP http://www.ncbi.nlm.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NCI-CGAP clone distribution
            information can be found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=NO.
FEATURES
     source
            Location/Qualifiers
            1..536
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2723471"
            /clone_lib="NCI_CGAP_Sub3"
            /lab_host="DH10B (Life Technologies)"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; The
            NCI_CGAP_Sub3 library is a subtracted library derived from
            the NCI_CGAP_Sub1 library, which is a subtracted library
            derived from BI. BI constitutes a mixture of 21
            normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4,
            NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
            NCI_CGAP_Co16, NCI_CGAP_Kids, NCI_CGAP_Kid12,
            NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
            NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CL11, NCI_CGAP_Le12,
            NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
            NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
            NCI_CGAP_Brn25. These 21 libraries were pooled and a
            single-stranded DNA preparation of the resulting mixture
            was used as a tracer in a subtractive hybridization with
            a driver whose composition is detailed below:
            NCI_CGAP_Kid3 pool 1    LLAM 3333-3337, 3682-3683,
            3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775
            , 1500552-1502855); NCI_CGAP_Kids pool 1   LLAM 3338-3342
            , 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,
            1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
            LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
            1520904-1522439); NCI_CGAP_GC4 pool 1    LLAM 3164-3167,
            3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
            1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
            985608-986759, 1101192-1101959, 1217928-1220615);
            NCI_CGAP_Co10 pool 1    LLAM 2644-2653, 2871-2872 (IMAGE
            CloneIDs 1057416-1061255, 1144584-1145351). Subtraction
            was performed as previously described [Bonaldo, Lennon &
            Soares (1996): Normalization and Subtraction: Two
            Approaches To Facilitate Gene Discovery. Genome Research
            6, 791-806.
            TAG_LIB=NCI_CGAP_Lu5
            TAG_TISSUE=lung
            TAG_SEQ=CAAC"
BASE COUNT       129 a      143 c      146 g      118 t
ORIGIN

Query Match        17.0%;   Score 421;  DB 113;  Length 536;
Best Local Similarity  99.8%;  Pred. No. 5.2e-211;
Matches   471;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     91 cggcgccgcaagagaccgctgcgcacgcgaggagaagcgcggagacgtcggg 150
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGCGCCGCAAGAGACCCGTGCGCACCTGTGCGCACCTGCAGGAGAAGCGCGGAGACCGTCGGG 60

Qy    151 tgctccggcgggcccaaacaccgtgtacctgcaggtgtggcagagccgggactcg 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    61 TGCTCCGCGCGGCTCTACGTTCACCTGCCTGGGTGGTTGCAGCGGGGGACTCG 120

Db   211 ggcgccgcggctctacgtcttccgagttcaaccggtatcctcttcaactgtgtggagaagc 270
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCCGCGGCGCCTPACGTCTTCCCGAGTTCAACCGTATCTGTGGAGAAGGC 180

Qy   271 gttcaggagactcatgcaggagcacaagttaaagttgtcgcctggacaacatattcctg 330
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTTCAGAGACTCATGCAGGAGCACAAGTTAAAGGTGCTGCCTGGACAACATATTCCTG 240

Qy   331 acacgaatgcactggttctaatgtgtgggggcttaagtggaatgattccttactttaaggaa 390
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ACACGAATGCACTGGTTCTAATGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAA 300

Qy   391 acgggcttccaaagtgtgtactttctggacctccacacactggaaaaataccttcgaagca 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACCGGGCTTCCAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTTCGAAGCA 360

Qy   451 atcaaaaatatttctcgtccattgaaaagaaatagaaactggctctgtcggggcccactctgcc 510
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ATCAAAAATATTTTCTGGTCCATTGAAAAGAATAGAAACTGGCTCTGTCGGGGCCCCACTCTGCC 420

Qy   511 ccggaatacgagggatgaaacccatgacagtttaccagatatccccatacacacagtg 562
         ||||||||||||||||||||||||||||||||||||||||
Db   421 CCAGAATACGAGGAGATGAAACCATGACAGTTTACCAGATACCCCCATACACAGTG 472
```

```
RESULT  36
LOCUS        BG327066         452 bp      mRNA            EST         27-FEB-2001
DEFINITION   602426274F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4564170 5',
             mRNA sequence.
ACCESSION    BG327066
VERSION      BG327066.1  GI:13133503
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 452)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgaps-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1279 row: c column: 19
             High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4564170"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      93 a    132 c    144 g    83 t
ORIGIN
```

```
Query Match     16.4%;    Score 406;  DB 152;  Length 452;
Best Local Similarity   100.0%;  Pred. No. 4.4e-203;
Matches  406;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    36 cggacgcaccatgtcgcagggacgcaccatatcgcaggcacccgccccgccgggagggcc 95
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    47 CGGACGCACCATGTCGCAGGGACGCACCATATCGCAGGCACCCGCCCCGCCGGGAGGGCC 106

Qy    96 gcgcaaggacccgtctgcggcacctgcgcacgcgagagagaagcagcgggacgtcgggggtgtc 155
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   107 GCGCAAGGACCCGTCTGCGGCACCTGCGCACGCGAGAGAGGCGGGACGTCGGGGGTCTC 166

Qy   156 cggcggcccaaacaccgtgtacctgcaggtgtgtggcagcgggtagccgggacctcgggcgc 215
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   167 CGGCGGCCCAAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGC 226

Qy   216 cgcgctctacgttcttccggagttcaaccgggtatctctcttcaactgtggagagaagcggttca 275
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   227 CGCGCTCTACGTTCTTCCGGAGTTCAACCGGGTATCTCTCTTCAACTGTGGAGAAGCGGTTCA 286

Qy   276 gagactcatgcaggagcacaagttaaagttgctgcctggacaacaatattcctgacacg 335
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   287 GAGACTCATGCAGGAGCACAAGTTAAAGGTTGCTGCCTGGACAACAATATTCCTGACACG 346

Qy   336 aatgcactggttctaatgtgtggggggcttaagtggaatgattcttacttttaaggaaacacgg 395
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   347 AATGCACTGGTTCTAATGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAAACCGG 406

Qy   396 gcttccaaagtgtgtactttctgggacctccacaactgggaaaaatac 441
         |||||||||||||||||||||||||||||||||||||||||||||
Db   407 GCTTCCAAAGTGTGTACTTTCTGGGACCTCCACAACTGGAAAAATAC 452
```

```
RESULT  37
LOCUS        BF309436         457 bp      mRNA            EST         21-NOV-2000
DEFINITION   601892128F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137654 5',
             mRNA sequence.
ACCESSION    BF309436
VERSION      BF309436.1  GI:11256824
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 457)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgaps-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution Information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM1045 row: h column: 07
             High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1..457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4137654"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
```

RESULT 38

LOCUS   AW304130   404 bp   mRNA   EST   18-JAN-2000
DEFINITION xs13e05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2769536 3'
            similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD PROTEIN IN
            SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AW304130
VERSION   AW304130.1 GI:6713819
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 404)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

FEATURES   Location/Qualifiers
        source   1..404
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2769536"
                        /clone_lib="NCI_CGAP_Kid11"
                        /lab_host="DH10B"
                        /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                        a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                        Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                        prepared, and ss circles were made in vitro. Following HAP
                        purification, this DNA was used as tracer in a subtractive
                        hybridization reaction. The driver was PCR-amplified cDNAs
                        from a pool of 5,000 clones made from the same library
                        (cloneIDs 1322376-1323911, 1456007-1456775, and
                        1500552-1502855). Subtraction by Bento Soares and M.
                        Fatima Bonaldo. "
BASE COUNT   133 a   145 c   85 g   94 t
ORIGIN

Query Match   16.3%; Score 405; DB 147; Length 457;
Best Local Similarity   100.0%; Pred. No. 1.5e-202;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 atgtgggcgcttgctcgctcgctcgctcgctcggcgacgcaccatgtcgcaggagacc 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15 ATGTGGGCGCTTGCTCGCTCGCTCGCTCGCTCGGCGACGCACCATGTCGCAGGGACGC 74

Qy   61 accatatcgcaggcaccccgccccgcccgcccgagcggccgccgcaaggaccgctgggggcacctg 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 ACCATATCGCAGGCACCCCGCCCCGCCCGCCCGAGCGGCCGCCGCAAGGACCCCCTGGGGCACCTG 134

Qy   121 cgcacgcagagaagcgcggggaccgtcggggtgtcgctccggcggggcccaaacaccgtgtacctg 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135 CGCACGCAGAGAAGCGCGGGACCGTCGGGGTGTCGCTCCGGCGGGCCCAAACACCGTGTACCTG 194

Qy   181 caggtggtgcaagcgggtagcgggactcgggcgccgcgcctcacgtcttctccgagttc 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195 CAGGTGGTGGCAACGGGGTAGCGGGACTCGGGCGCCGCGCCTCACGTCTTCTCCGAGTTC 254

Qy   241 aaccggtatcctcaacgttcaactcatgcaggaagcaacaagtta 300
         |||||||||||||||||||||||||||||||||||||||
Db   255 AACCGGTATCTCTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCAACAAGTTA 314

Qy   301 aaggtgttcgtcctggacaacatattcctgacacgaatgcactggtcta 360
         |||||||||||||||||||||||||||||||||||||||
Db   315 AAGGTTGCTCGCCCTGGACAACATATTCCTGACACTGGTTCTTAATGTTGTGGGGGC 374

Qy   361 ttaagtggaatggattcttacttttaäaggaaaccgggcttccaaag 405
         ||||||||||||||||||||||||||||||||||||||
Db   375 TTAAGTGGAATGGATTCTTTACTTTAAAGGAAAACCGGGCTTCCAAAG 419

RESULT 39

LOCUS   AA291670   474 bp   mRNA   EST   16-MAY-1997
DEFINITION zt37d04.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
            IMAGE:724519 3' similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8
            KD PROTEIN IN SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AA291670
VERSION   AA291670.1 GI:1939648
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 474)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
TITLE   WashU-Merck EST Project 1997
JOURNAL   Unpublished (1997)

        source   1..404
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:2769536"
                        /clone_lib="NCI_CGAP_Kid11"

BASE COUNT   96 a   107 c   113 g   88 t
ORIGIN

Query Match   16.3%; Score 404; DB 114; Length 404;
Best Local Similarity   100.0%; Pred. No. 5e-202;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   110 tgcggcacctgcgcacgcagagaagcgcgggaccgtcggggtgtcgctccggcggggcccaaaca 169
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TGCGGCACCTGCGCACGCAGAGAAGCGCGGGACCGTCGGGGTGTCGCTCCGGCGGGCCCAAACA 60

Qy   170 ccgtgtacctgcaggtggtgcaagcgggtagcgggactcgggcgccgcgcctcacgtct 229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCGTGTACCTGCAGGTGGTGGCAACGGGGTAGCGGGACTCGGGCGCCGCGCCTCACGTCT 120

Qy   230 tctccgagttcaaccggtatcctcaactgtggagaaggcgttcagagactcatgcagg 289
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCTCCGAGTTCAACCGGTATCTCTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGG 180

Qy   290 agcacaagttaaagttgctcgccctggacaacatattcctgacacgaatgcactggtcta 349
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGCACAAGTTAAAGGTTGCTCGCCCTGGACAACATATTCCTGACACTGGTTCATGCAGG 240

Qy   350 atgttggggcttaagtggaatgattcttacttttaaaggaaaccgggcttccaaagtgtg 409
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATGTTGGGGCGCTTAAGTGGAATGATTCTTTACTTTAAAGGAAAACCGGGCTTCCAAAGTGTG 300

Qy   410 tactttctggacctccacacaactggaaaaatacctcgaagcaatcaaaatatttctggtc 469
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAAATATTTTCTGGTC 360

Qy   470 cattgaaaggaatagaactgggctgtcggggccccaactctgcccca 513
         |||||||||||||||||||||||||||||||||||||||||||||
Db   361 CATTGAAAGGAATAGAACTGGGCTGTCGGGGCCCCCAACTCTGCCCCA 404

COMMENT   Contact: Wilson RK
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           This clone is available royalty-free through LLNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Possible reversed clone: similarity on wrong strand
           Possible reversed clone: polyT not found
           Seq_primer: -41m13 fwd. ET from Amersham
           High quality sequence stop: 465.
FEATURES   Location/Qualifiers
     source   1. .474
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /clone="IMAGE:72419"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGGCGGCCGCGGGTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
BASE COUNT   128 a   120 c   129 g   97 t
ORIGIN

Query Match
Best Local Similarity   100.0%;   Score 402;   DB 4;   Length 474;
Matches 402;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   201 ccgggactgggcgcgcccgcgtctacgtctctccgagttcaaacggttattcttcaactg 260
          ||||||||| ||||||| ||||||||||||| ||||||||||||||||||||||||||||||
Db   73 CCGGGACTCGGGCGCCGGCCTCTACGTCTTCTCCGAGTTCAAACGGTTATTCTTCAACTG 132

Qy   261 tggagaaggccgttcagagaactcatgcaggagcacaagttaaagttgttctcgcctggacaa 320
          ||||||||||||||||||||| ||||| ||||||||||||||||| ||||| |||||||||||
Db   133 TGGAGAAGGCCGTTCAGAGACTCATCCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAA 192

Qy   321 catattcctgacacgaatgcactggtctaatgtgttgggggcttaagtgaatgattcttac 380
          |||||||||||||||||||||||||||||||| |||||||| |||||| ||||||||||||
Db   193 CATATTCCTGACACGAATGCACTGGTCTAATGTTTGGGGGGCCTTAAGTGGAATGATTCTTAC 252

Qy   381 tttaaaggaaaccgggcttcccaaagtgttgtcacttcctggacctccacaactggaaaata 440
          |||||||| |||||||||||||| ||||||||| |||||||||||||||||||||||||||
Db   253 TTTAAAGGAAAACCGGGCTTCCAAAGTGTGTTGTTACTTTCTGGACCTCCACAACTGGAAAATA 312

Qy   441 cctcgaagcagaatcaaaaatatttcttcggtccattgaaaggaatagaaactggctgcgcggcc 500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   313 CCTCGAAGCAATCAAAAATATTTTCTTCGGTCCATTGAAAGGAATAGAAACTGGCTGTGCGGCC 372

Qy   501 ccactctgcccccagaatacgagggatgaaaccatgacagtttaccagaatccccaatacacag 560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 CCACTCTGCCCCCAGAATACGAGGGATGAAAACCATGACAGTTTACCAGAATCCCCAATACACAG 432

Qy   561 tgaacagagagaggggaaaggcaccaccaccatggcaggagtccaga 602
          |||||||||||||||| ||||||||||||||||||||||||||||||
Db   433 TGAACAGAGAGAGGGGAAAGCACCACCAACCATGGCAGGAGTCCAGA 474

VERSION   BF525432.1   GI:11612793
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 947)
  AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: David N. Louis, M.D.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           High quality sequence stop: 695.
           Plate: LLAM9783   row: a   column: 06
FEATURES   Location/Qualifiers
     source   1. .947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4212365"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   224 a   257 c   289 g   177 t
ORIGIN

Query Match
Best Local Similarity   99.8%;   Score 400;   DB 150;   Length 947;
Matches 520;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   1;

Qy   16 gcaggaatacagagggagtgcggagagggacggccccagcgccccagccagcagagaaaagaagtcagta 79
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 GCAGGAGTACAGAGGGAGGAGTGCGGCAGGAGCACGAGCGCCCAGCCGCCAGCAGAGAAAAGAAGTCAGTA 1439

Qy   1440 cccagaaatcatcctcctggaacacagggtctgccatcccgatgaagagattcgaaatgtcag 1499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCCAGAAATCATCTTCCTTGGAACACAGGGTCTGCCAATCCCGATGAAGAAGATTCGAAATGTCAG 120

Qy   1500 tgccacaacttgtcaaacataagccccgacacgtctcctgctactggactgtggttgaggcac 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGCCACACTTGTCACAACATAAGCCCCGACAACGTCTCGCTACTGGACTGTGGTTGAGGGCAC 180

Qy   1560 atttgggcagctgtgccgtcattacggagaccaagtgatcagggtccatgagttgcccaccctggc 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGAGGTCCATGAGTTGCCCACCCTGGC 240

Qy   1620 tgctgtgttttgtgtccacctgcacgcacgccaagatcaccccgaaagccttgccaagtatcttgct 1679
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGCTGTG-TTGTGTCCACCTGCACGCAGATCACCACGCAGCCTTGCCAAGTATCTTGCT 299

Qy   1680 gcaggagaaacgcgcccttggcatcttctttgggaaaagcggcttcaccccttgctggttggtggc 1739
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   300 GCAGGAGAGAACCGCGCCCTTGGCATCTTTGGGAAAAGCCGCTTCACCCTTTGCTGGTTGGTTGC 359

Qy   1740 cccaaccagctcaaagcctgcctgacccagctaccacacaaccagtgccaggaggtcctgca 1799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   360 CCCCAACCAGCTCAAAGCCTGCCTGACCCAGCTACCACAACAACCAGTGCCAGGAGGTCCTGCA 419

Qy   1800 ccacatcaagtatgattcctgccaaatgccttcaggaaagggctgagaatctccagctcctgc 1859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   420 CCACATCAGTATGATATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGAATCTCCAGTCCTGC 479

RESULT 40
LOCUS   BF525432   947 bp   mRNA   EST   11-DEC-2000
DEFINITION 602069517F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4212365
            5', mRNA sequence.
ACCESSION   BF525432

RESULT 41

```
Qy    1860  agtcggaaagaaattgatcagttcgctgttcggaacatgtgatt 1900
            ||||||||||||||||||||||||||||||||||||||||||||
Db     480  AGTCGGAAAGAATTGATCAGTTCGCTGTTCGGAACATGTGATT 520
```

```
Db  BE747163
LOCUS       BE747163      992 bp    mRNA                    EST    15-SEP-2000
DEFINITION  601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
            mRNA sequence.
ACCESSION   BE747163
VERSION     BE747163.1  GI:10161155
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 992)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM524 row: 1 column: 14
            High quality sequence stop: 781.
FEATURES             Location/Qualifiers
     source          1..992
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3838237"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      225 a    287 c    313 g    167 t
ORIGIN
```

```
Qy    1082  ttgggcctgaccaccagcaacttggtctcctgaatagagaactgtgcctcagttcacaacctc 1141
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      90  TTGGGCCTGACCACCAGCAACTTGGTCTCCTGAATGAGAGAACTGTGCCTCAGTTCACAACTTC 149

Qy    1142  gcagccacaaaaattcaaacccagctcaaactcatccacccggacatctcccccctgctca 1201
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150  GCAGCCACAAGANTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCTGCTCA 209

Qy    1202  ccagttccgctgtaagaaggagggcccacacctcagttgtgccatggtgttcaggttgaat 1261
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210  CCAGTTCCGCTGTTAAGAAGGAGGGCCCACACCTCAGTTGTGCCATGGTTCAGGTGAAT 269

Qy    1262  gcctcctcaagtaccagctccgtccccaggaggaggggagtagtcaggaggggatagccaatt 1321
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270  GCCTCCTCAAGTACCAGCTCCGTCCCCAGGAGGAGGTGGCAGAGGGATGCCATTATTACTT 329

Qy    1322  gcaatcctgaggaattcaatagttgaggcgctgcagcttcccaacttccagcagagcgtgc 1381
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     270  GCCTCCTCAAGTACCAGCTCCGTCCCCAGGAGGAGGTGGCAGAGGGATGCCATTATTACTT 1381
```

Query Match   16.1%;   Score 400;   DB 139;   Length 992;
Best Local Similarity 99.7%;   Pred. No. 6.9e-200;
Matches 570;   Conservative 0;   Mismatches 1;   Indels 1;   Gaps 1;

RESULT 42

```
Db     330  GCAATCCTGAGGAATTCATTAGTTGAGGCGGTCGCAGCTTCCAACTTCCAGCAGACGTGC 389
```

```
Qy    1382  aggagtacaggagaagtcgccagaagacgcgccaagccccagcccaggaagaagtcagtacc 1441
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     390  AGGAGTACAGGAGGAGAGTCGCCAGAGCC-CCAGCGCCAGCAGAGGAAGAAGTCAGTACC 448

Qy    1442  cagaaatcatcttcctggaacagggtgtgacatcccgatcgaaagatccgaaatgtcaagtg 1501
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449  CAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATCGAAAGATTCGAAATGTCAGTG 508

Qy    1502  ccacacttgtcaaacatacagcccccagacacgtctctgctactggactgtggtgaggcacat 1561
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     509  CCACACTTGTCAAACATAAGCCCGACACACGTCTTTGCTACTGGACTGTGTGGGGCACAT 568

Qy    1562  ttgggcagctgtgtgcgtgcattacggagaccaaggtggacaggtcctgggccaccctggctg 1621
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     569  TTGGGCAGCTGTGTGCCGTGCATTACGGAGACCAAGGTGGACAGGTCCTGGGCACCCTGGCTG 628

Qy    1622  ctgtgtgtttgtgtccacctgcaacgcagagatcac 1653
            ||||||||||||||||||||||||||||||||||||
Db     629  CTGTGTTTGTGTCCACCTGCAACGCAGATCAC 660
```

```
Db  BE386924
LOCUS       BE386924      627 bp    mRNA                    EST    21-JUL-2000
DEFINITION  601274815F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3615939 5',
            mRNA sequence.
ACCESSION   BE386924
VERSION     BE386924.1  GI:9332387
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM280 row: n column: 04
            High quality sequence start: 18
            High quality sequence stop: 611.
FEATURES             Location/Qualifiers
     source          1..627
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3615939"
                     /clone_lib="NIH_MGC_20"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      147 a    175 c    199 g    106 t
ORIGIN
```

Query Match   15.6%;   Score 387;   DB 167;   Length 627;
Best Local Similarity 100.0%;   Pred. No. 5e-193;
Matches 387;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy   2095 gaaaagaacaacagcacaaacgtcccaaagccaatcagcgtgggagatgcggatgaacgcggag 2154
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16 GAAAAGAACAACAGCACAACGTCCCAAAGCCAATCACCGTGGGAGATGCGGATGAACGCGGAG 75

Qy   2155 ttcattatgctgaaccacttcagccagcgctatgccaaggtcccccctcttcagccccaac 2214
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 TTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCCTCTTCAGCCCCAAC 135

Qy   2215 ttcagcgagaaagtggggagggttgccttgaccacatgaaggtctgctttgagacttttcca 2274
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGGACTTTTCCA 195

Qy   2275 acaatgcccaagctgattccccactgaaagcccctgtttgctgggacatcgaggagatg 2334
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 ACAATGCCCAAGCTGATTCCCCACTGAAAGCCCCTGTTTGCTGGGACATCGAGGAGATG 255

Qy   2335 gaggagacgcaggagaagcgggagagctgcgcggggcccaagggggag 2394
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 GAGGAGACGCAGGAGAAGCGGGAGAGCTGCGCGGGGCCCAAGGGGGAG 315

Qy   2395 ctggcaggcggcctggggagccttcagcaggaagaggaggagagcca 2454
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    316 CTGGCAGGCGGCGCCTGGAGGATGGGGAGCCTTCAGCAGGAAGAGGAGGAGAGCCA 375

Qy   2455 caggccaagaaggtcagagcccagtga 2481
          |||||||||||||||||||||||||||
Db    376 CAGGCCAAGAAGGTCAGAGCCCAGTGA 402
```

```
RESULT 43
LOCUS       BE250309          574 bp    mRNA           EST       13-JUL-2000
DEFINITION  6009434S5F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960077 5',
            mRNA sequence.
ACCESSION   BE250309
VERSION     BE250309.1  GI:9120418
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM52  row: n  column:14
            High quality sequence stop: 571.
FEATURES             Location/Qualifiers
     source          1..574
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2960077"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT    124 a    158 c    167 g    119 t    6 others
```

```
ORIGIN

Query Match   15.5%;   Score 384;   DB 165;   Length 574;
Best Local Similarity 99.8%;  Pred. No. 1.9e-191;
Matches  434;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1574 gccgtcattacggagaccaaggtggacaggtcctgggcaacctggctgctgtttgtgt 1633
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCGTCATTACGGAGACCAAGGTGGACAGGTCCTGGGCAACCTGGCTGCTGTTTGTGT 60

Qy   1634 cccacctgcacggagatcaccacacacggggcttgccaagtatcttgtctgcaagagaaacgcg 1693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCACCTGCACGGAGATCACCACACACGGGCTTGCCAAGTATCTTGTCTGCAAGAGAAACGCG 120

Qy   1694 ccttggcatcttgggaaagccgcttcaccccttgctggtgtggttgccccaccagctca 1753
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCTTGGCATCTTTGGGAAAGCCGCTTCACCCCTTTGCTGGTGTGGTTGCCCCACCAGCTCA 180

Qy   1754 aagcctggcctccagcagtaccacaacaacaagtgccaggaggtcctgcaaccacatcagtatga 1813
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAGCCTGGCCTCCAGCAGTACCACAACAACCAAGTGCCAGGAGGTCCTGCAACCACATCAGTATGA 240

Qy   1814 ttcctgccaaatgccttcaaggaaggggtgagatctccagtctgcagtggaaagattga 1873
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTCCTGCCAAATGCCTTCAAGGAAGGGGTGAGATCTCCAGTCTGCAGTGGAAAGATTGA 300

Qy   1874 tcagttcgctgttgcgcaacatgtgattcttggaagagttcagacctgtctggttgcggcact 1933
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCAGTTCGCTGTTGCGCAACGTGTGATTTTGGAAGAGTTTCAGACCTGTCTGGTTGCGGCACT 360

Qy   1934 gcaaagcatcgcgtttggggctgtctggtgtgccgcctggtgtcacaccctctggcggaaagtggctga 1993
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAAAGCATCGCGTTTGGGGCTGTCTGGTGTGCCGCCTGGTGTCACACCCTCTGGCGGAAAGTGGCTCTGAT 420

Qy   1994 gggaacacatgccct 2008
          |||||||||||||||
Db    421 GGGAACACATGCCCT 435
```

```
RESULT 44
LOCUS       BF732987          397 bp    mRNA           EST       30-MAR-2001
DEFINITION  nae07b01.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3434448 3',
            similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF732987
VERSION     BF732987.1  GI:12058062
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. CDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 354.
FEATURES             Location/Qualifiers
     source          1..397
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                    /clone="IMAGE:3434448"
                    /clone_lib="NCI_CGAP_Ov18"
                    /tissue_type="fibrotheoma"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                    strand cDNA was primed with a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCGCCGCGGACATTTTTTTTTTTTTT 3'];
                    double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not
                    I and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT     95 a    101 c    115 g     86 t
ORIGIN

Query Match       15.4%;   Score 383;   DB 168;   Length 397;
Best Local Similarity  100.0%;   Pred. No. 6.4e-191;
Matches 383;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   108 gctggcgcacctgcgcacgcgaggagaacgcggagcggggagcggagcggggagggcctctacgt 167
         ||||||||||||||||||| |||||||||  |||||| |||||||  |||||||||||||||||
Db     1 GCTGGCGCACCTGCGCACCGCAGGAAGCGCGGACCGTCGGGGGGTTGCTCCGGCCCAAA 60

Qy   168 caccgtgtaccctgcaggtgtggtcggaacggggtagccggggcgcgcggccgcgctctacgt 227
         ||||||||||||| ||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    61 CACCGTGTACCTGCAGGTGTGGTCGGGGACTCGGCGCGGGGGCGCGGCCGCGCTCTACGT 120

Qy   228 cttctccgagttcaaccggtatctcttcaactgtggagaagcgttcagagactcatgca 287
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   121 CTTCTCCGAGTTCAACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCA 180

Qy   288 gggacacaagttaaagttgctcggcctggacacaacatatcctgacacgaatgcactggtc 347
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGGACACAAGTTAAAGGTTCTCGGCCTGGACACAACATATTCCTGACACATATTCCTGAC 240

Qy   348 taatgttggggggttaagtggaatgaatgatttacttttaaagaaacggggcttcaaagtg 407
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TAAGTGTTGGGGGGTTAAGTGGAATGAATGATTTACTTTTAAAGGAAACCGGGCTTCAAAGTG 300

Qy   408 tgtacttttctgtgacctccacacaactggaaaaaatacctcgaagcaaatcaaaatatttctgg 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TGTACTTTTCTGTGACCTCCACAACTGGAAAAAATACCTCGAAGCAATCAAAATATTTTCTGG 360

Qy   468 tccattgaaaggaatagaactgg 490
         |||||||||||||||||||||||
Db   361 TCCATTGAAAGGAATAGAACTGG 383
```

```
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9499  row: n  column: 01
            High quality sequence stop: 631.

FEATURES             Location/qualifiers
     source          1..878
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4183488"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.57 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT    208 a    250 c    230 g    190 t
ORIGIN

Query Match       15.4%;   Score 382;   DB 147;   Length 878;
Best Local Similarity   99.8%;   Pred. No. 2.3e-190;
Matches 502;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy  1082 ttgggcctgacacccagccacttggtcctgaatgagaaactgttgccagtcaacaacctc 1141
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 TTGGGCCTGACACCCAGCCACTTGGTCCTGAATGAGAAACTGTTGCCAGTCAACAACTTC 176

Qy  1142 gcagccacaagagttcaaaccagcagcttcaaacctcatccaccacccggacacatcttccccctgctca 1201
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 GCAGCCACAAGATTCAAAACCAGCCTCAAACCTCATCCACCACCCGGACACATCTTCCCCTGCTCA 236

Qy  1202 ccagttccgctgtaagaaggagggcccaccacctcagttgcccatgttcaggtgaat 1261
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CCAGTTCCGCTGTAAGAAGGAGGGCCCACCACCTCAGTGTGCCCATGGTTCAGGGTGAAT 296

Qy  1262 gcctcctcaagtaccaagctccgtcccagagggggagttggcagaggggatgccattattactt 1321
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 GCCTCCTCAAGTACCAGCTCCGTCCCAGAGGGGAGGAGATGCCAATTATTATTACTT 356

Qy  1322 gcaatcctgaggaattcaatagttgagcgtgcagcttccaaacttccacagagacgtgc 1381
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 GCAATCCTGAGGAATTCATTAGTTGAGGCGTGCAGCTTCCCAACTTCCACAGAGACGTGC 416

Qy  1382 aggagtacaggagagtcgcgaggacggacccagccagcccacagcagaagaagtcagtacc 1441
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   417 AGGAGTACAGGAGGAGTCGCCGCAGGACGG-CCAGCCCCAGCGAGAGAAAAGAAGTCAGTACC 475

Qy  1442 cagaaatcatcttcctggaacagggtctgccatccgatgaagattcgaaatgtcagtg 1501
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   476 CAGAAAATCATCTTCCTTGGAACAGGGTCTGCCATCCGATGAAGATTCGAAATGTCAGTG 535

Qy  1502 ccacacttgtcaaacataagcccgacacgtctctcgtcctactggaactgtgttgtaggccacat 1561
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   536 CCACACTTGTCAAACATAAGCCCCGACACGTCTCTCGTCCTACTGGACTGTGGTGTGAGGGGCACAT 595

Qy  1562 ttgggcagctgtgccgtcattac 1584
         |||||||||||||||||||||||
Db   596 TTGGGCAGCTGTGCCGTCATTAC 618
```

Search completed: October 28, 2001, 20:05:22
Job time: 7221 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: October 28, 2001, 20:05:27 ; Search time 1393.91 Seconds
                                              (without alignments)
                                              11107.759 Million cell updates/sec

Title: SEQIINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacaagattcaaa.........aagccatacgcgtgggatg 1001

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Post-processing: Minimum Match 0%
Maximum DB seq length: 0
Minimum DB seq length: 2000000000

Database :   GenEmbl:*
                 Maximum Match 100%
                 Listing first 45 summaries

```
 1: gb_ba1:*            44: em_ov:*
 2: gb_ba2:*            45: em_pat:*
 3: gb_ba3:*            46: em_ph:*
 4: gb_in1:*            47: em_pl:*
 5: gb_in2:*            48: em_ro:*
 6: gb_in3:*            49: em_sts:*
 7: gb_om:*             50: em_sy:*
 8: gb_ov:*             51: em_un:*
 9: gb_pat1:*           52: em_vi:*
10: gb_pat2:*           53: gb_sts1:*
11: gb_ph:*             54: gb_sts2:*
12: gb_pl1:*            55: gb_sts3:*
13: gb_pl2:*            56: gb_sy:*
14: gb_pl3:*            57: gb_un:*
15: gb_pl4:*            58: gb_vil:*
16: gb_ba1:*            59: gb_vi2:*
17: em_ba2:*            60: gb_htg1:*
18: em_fun:*            61: gb_htg2:*
19: em_htg_hum:*        62: gb_htg3:*
20: em_htgo_inv:*       63: gb_htg4:*
21: em_htgo_rod:*       64: gb_htg5:*
22: em_htg_hum1:*       65: gb_htg6:*
23: em_htg_hum2:*       66: gb_htg7:*
24: em_htg_hum3:*       67: gb_htg8:*
25: em_htg_hum4:*       68: gb_htg9:*
26: em_htg_hum5:*       69: gb_htg10:*
27: em_htg_hum6:*       70: gb_htg11:*
28: em_htg_hum7:*       71: gb_htg12:*
29: em_htg_hum8:*       72: gb_htg13:*
30: em_htg_inv1:*       73: gb_htg14:*
31: em_htg_inv2:*       74: gb_htg15:*
32: em_htg_other:*      75: gb_htg16:*
33: em_htg_rod:*        76: gb_htg17:*
34: em_hum1:*           77: gb_htg18:*
35: em_hum2:*           78: gb_htg19:*
36: em_hum3:*           79: gb_htg20:*
37: em_hum4:*           80: gb_htg21:*
38: em_hum5:*           81: gb_htg22:*
39: em_hum6:*           82: gb_htg23:*
40: em_hum7:*           33: gb_htg24:*
41: em_in:*             14: gb_htg25:*
42: em_om:*              5: gb_pr1:*
43: em_or:*              6: gb_pr2:*
                         7: gb_pr3:*
                        78: gb_pr4:*
                        79: gb_pr5:*
                        89: gb_pr6:*
                        90: gb_pr7:*
                        91: gb_pr8:*
                        92: gb_pr9:*
                        93: gb_pr9:*
                        94: gb_ro1:*
                        95: gb_ro2:*
                        96: gb_in4:*
                        97: gb_pr10:*
                        98: em_ba3:*
```


(handwritten notes)
Insertion of seq 10 NO1
std of seq
search of
1000 residues
marketing
insertion
insertion
site

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	989	98.8	2908	89	AF304370	AF304370 Homo sapi
2	987.4	98.6	2997	91	BC001939	BC001939 Homo sapi
3	987.4	98.6	3006	91	BC004158	BC004158 Homo sapi
4	973	97.2	2908	89	AF308698	AF308698 Pan trogl
5	971.4	97.0	2976	89	AK001392	AK001392 Homo sapi
6	965	96.4	2893	89	AF308694	AF308694 Gorilla g
7	728.8	72.8	2712	94	AF308696	AF308696 Mus muscu
8	137	13.7	118788	85	AC005277	AC005277 Homo sapi

RESULT 1

9	135.2	13.5	740	89	AF3047152	AF304369	Homo sapi
10	103	10.3	34593	94	AF348157	AF348157	Mus muscu
11	71.4	7.1	31201	15	SPAC1D4	Z69239	S.pombe chr
12	70.6	7.1	15873	64	AC017383	AC017383	Drosophil
13	70.6	7.1	15682	4	AC007417	AC007417	Drosophil
14	70.6	7.1	190574	60	AC007352	AC007352	Drosophil
15	70.6	7.1	261846	5	AE003830	AE003830	Drosophil
16	69	6.9	2690	5	AF215894	AF215894	Drosophil
17	51.6	5.2	32329	15	SPBC3D6	Z95620	S.pombe chr
18	48	4.8	11836	12	AE000936	AE000936	Methanoba
19	47.4	4.7	952	53	CNS079WW	AL435846	T7 end of
20	46.8	4.7	143209	13	AP001550	AP001550	Oryza sat
21	45.4	4.5	19158	13	AF188714	AF188714	Emericell
22	43.2	4.3	46335	12	AC022354	AC022354	Arabidops
c 23	42	4.2	12026	1	AE005109	AE005109	Halobacte
24	41.4	4.1	75289	12	AB023046	AB023046	Arabidops
25	39.4	3.9	304	92	HS12H2R	Z64511	H.sapiens C
26	38.4	3.8	23533	14	AE001039	AE001039	Archaeogl
27	38	3.8	1245	11	CREPHOSRI	M36123	C.reinhardt
28	38	3.8	2652	13	AF228914	AF228914	Chlamydom
c 29	38	3.8	40360	3	MTV049	AL022021	Mycobacte
c 30	37.8	3.8	167191	60	AC008875	AC008875	Homo sapi
c 31	37.8	3.8	193567	60	AC008945	AC008945	Homo sapi
c 32	37.8	3.8	236029	65	AC019232	AC019232	Homo sapi
c 33	37.6	3.8	723	54	G28337	G28337	human STS S
c 34	37.6	3.8	724	97	HUMRRP	D29833	Homo sapien
c 35	37.6	3.8	7618	85	AB031740	AB031740	Homo sapi
c 36	37.6	3.8	17323	68	AC024526	AC024526	Homo sapi
c 37	37.2	3.7	61552	77	AC087284	AC087284	Homo sapi
c 38	37.2	3.7	183097	71	AC027638	AC027638	Homo sapi
c 39	37	3.7	3685	15	SCYKR079C	Z28304	S.cerevisia
c 40	37	3.7	29861	90	CELW03D2	AF000298	Caenorhab
c 41	36.8	3.7	125673	90	AL356796	AL356796	Human DNA
c 42	36.8	3.7	286650	2	AP000995	AP000995	Thermopla
c 43	36.4	3.6	10732	10	E32986	E32986	Gene encodi
c 44	36.4	3.6	22846	6	CELE04A4	AF038611	Caenorhab
45	36.4	3.6	163691	73	AC067858	AC067858	Homo sapi

```
LOCUS       AF304370    2908 bp    mRNA    PRI    23-FEB-2001
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            (HPC2/ELAC2) mRNA, complete cds.
ACCESSION   AF304370
VERSION     AF304370.1  GI:10880932
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J.,
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED    11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     CDS             1..2481
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAG24441.1"
                     /db_xref="GI:10880933"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSCSGGPNTVILQVVAAGSRDSGAALIVFSEENRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWHWSNYGGLSGMILTLKETGLPKCVLSGPPQLEKLEAILKIFSGPLKG
                     IELAVRPHSAREYEDETMTVYQIPHSEQRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVAFICKLHLKRGNFLVLRAKEMGLPVGTRAIAPIIA
                     AVKDKSITHEGREILAEELCTPDPDGAAFVVVEQPDESFIOPICENATFQRYQKRAD
                     APVALVHMAPASVLVDSRYQQWMBERGPDTQHLVLNENCASVHNLRSHKTQTQLNLI
                     HPDIFPLLTSPRCKKEGPTLSVPMVGGECLLKYQLRPRREWQRDALITCNPEEFIVEA
                     LQLPNFQQSVQEYRSRSAQDGPAPAARKSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLDCGEGTFGQLCRHYGDQVDRVLGTIAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGRPLHPLLVAPNQLKAWLQQYHNCQQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLIRTCDLEEFQTCLVRHCKHAFGCALYISFSGWKVVVSGDTMPCEALVRMGGHATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKYPLFSDNFS
                     EKVGVAPDHMKVCFGDPPTMPKLIPPLKALPFAGDIEMEEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQOKRAHTEEPQAKKVRAQ"
```

ALIGNMENTS

```
Query Match    98.8%;   Best Local Similarity    99.9%;    Score 999;  DB 89;  Length 2908;
Matches 1000;  Conservative    0;  Mismatches  0;  Indels  1;  Gaps  1;    Pred. No. 1e-261;
```

BASE COUNT   702 a   793 c   826 g   587 t
ORIGIN

```
Qy     1 tcgcagccacaagaattcaaaaccccagtctcaaacctcacccaccccggacatctcccctgct 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1140 TCGCAGCCACAAGAATTCAAAAACCCCAGTCTCAAACCTCACCCACCCCGGACATCTCCCCTGCT 1199

Qy    61 caccagtttccgctgtaagagaagaggccccaccctcagtgtgcccatggttcaggtga 120
         |||||||||| |||||| |||||||| ||||||   |||||||||||||||||||||||
Db  1200 CACCAGTTTCCCCTGTGAAGAAGAGGAGGCCCCCAACCTCAGTGTGCCCATGGTTCAGGTGA 1259

Qy   121 atgcctcctcaagtaccagctccgtcccagaggaggaggatggcagaggagtgccattattac 180
         ||||||||||||||||||||||||||||||||||||||||||||||          |||||||
Db  1260 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGAGGAGGAGGATGGCCATTATTAC 1319

Qy   181 ttgcaatcctcaaggaattcatagttgaggccgtgcagcttcccacacttccgcagcagcgt 240
         ||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||| |||
Db  1320 TTGCAATCCTGAGGGAATTCATAGTTGAGGCCGTGCAGCTTCCCAACTTCCAGCAGAGCGT 1379

Qy   241 gcaggagtacaggaggaggtgcgcaggacggcgcccagcccagcagagaaaagaagtcagta 300
         |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  1380 GCAGGAGTACAGGAGGAGGTGCGCCAGGACGGCGCCCAGCCCAGCAGAGAAAAGAAGTCAGTA 1439

Qy   301 cccagaaatcatcttcctggaacagggtctgccatccgatgaaggatccgaaatgtcag 360
         |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  1440 CCCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCGATGAAGGATCCGAAATGTCAG 1499

Qy   361 tgccacacttgtcaacataagcccccgaacacgtctctgctactggaactgtgtgaggggcac 420
         ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  1500 TGCCACACTTGTCAACATAAGCCCCCGACACGTCTCTGCTACTGGAACTGTGTGAGGGGCAC 1559

Qy   421 atttggggcagctgtgtgccgtcattacggagacaccagttgacagaggggtcctggggcac 480
         |||||||||||||||||||||||||||||||||||  || |||| ||||| |||||||||||
Db  1560 ATTTGGGGCAGCTGTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGGCACCCTGGC 1619
```

```
Qy    481  tgctgtgttgttgtcccaactgggacgcagatcaccacacggggcttgccaagtatcttgc  540
           |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||||
Db   1620  TGCTGTGTTGTTGTGTCCCACCT-GCACGCAGATCACCACACGGGCTTGCCAGTATCTTGC  1678

Qy    541  tgcaggagaacgcgccttggcatccttgggaaagccgcttcacccttgctgtggttg  600
           |||||||||| ||||| ||||||||| |||||||||||||||||||||||||||||||||
Db   1679  TGCAGGAGAGAACGCGCCGCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTGCTGGTGGTTG  1738

Qy    601  ccccccaacagtcaaagcctctgcctcgacagcagtaccacaacagtgccagtcctgc  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1739  CCCCCCAACCAGTCAAAGCCTGGCTCCACCAGTACCACAACCACTGCCAGGAGGTCCTGC  1798

Qy    661  accacaatcaagtcatgtattcctgccaaatgccttccaggaagggggctgagatctccagtcctg  720
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1799  ACCACAATCAAGTCATGTATTCCCTGCCAAATGCCTTCCAGGAAGGGGGCTGAGATCTCCAGTCCTG  1858

Qy    721  cagtggaagaagattgaatcagttcgctgttgccgaacatgtgaattggaagaggtttcagacct  780
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1859  CAGTGGAAAGATTGATCAGTTCGCTGTTGCCGAACATGTGAATTGGAAGAGTTTCAGACCT  1918

Qy    781  gtcttggttgccggcaactgaacaagcatgcgtttggctgtgtgctggtgcacacctctggctgga  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1919  GTCTGGCTGCGGGCACTGCAAGCACATCCGTTTGGCTGTGTGGTCGCACCTCTGGCTGGA  1978

Qy    841  aagtggtctattccgggacacacatgtccctcgagctctgtgtccggatgggaaagatg  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1979  AAGTGGTCTATTCCGGGGACACACATGTCCCTCGGCTGTGTCCGGATGGGAAAGATG  2038

Qy    901  ccacctcctgatcatgaagccccctgggaagtggttggaggaagcagtggaaa  960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2039  CCACCTCCTGATCATGAAGCCACCCTGGAGAGGCACTGGAGAAGCACTGGAAA  2098

Qy    961  agacaacacaagcacaacgtccccaagccatccaagcgtgggatg  1001
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   2099  AGACACACAGCACAACGTCCCAAGCCATCAGCCGTGGGGATG  2139
```

```
RESULT  2
BC001939
LOCUS       BC001939             2997 bp    mRNA            PRI       16-MAR-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:4102, mRNA, complete cds.
ACCESSION   BC001939
VERSION     BC001939.1  GI:12804972
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2997)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomes Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
```

Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

```
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: i Column: 17.
FEATURES             Location/Qualifiers
     source          1..2997
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:4102 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             61..2541
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH01939.1"
                     /db_xref="GI:12804973"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDLRLRTRE
                     KRGPSGCCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILILKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYVQIPIHSEQRGHQPWQSEPRPLSRLSPERSSSDSELN
                     ENEPHLPHGVSQRGVRDSSLVVAFICKLHLKRGNFLVLRAKEMGLPVGTAAIAPIA
                     AVKDGKSITHEGREILAEELCTPPDGAAFVVVECDDESFIQPICENATFQRYQGRAD
                     APVALVVHAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVOGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAERKSQYPEIIELGTGSAIPMKIRNVSATLVNISP
                     DTSILLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLPSILLQRRAL
                     ASLGKPLHPLIVVAPNQLKAWLQQYHNCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAPGCALVHTSGWRKVVYSGDTMPCEALVRMGKDATL
                     LIHFATLEDGLEERAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSNFS
                     EKVGVAFDHHMKVCFGDFPTMPKLLIPPLKALFAGDIEEMERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRANTEEPQAKKVRAQ"
BASE COUNT      740 a    806 c    853 g    598 t
ORIGIN
```

```
Query Match        98.6%;  Score 987.4;  DB 91;  Length 2997;
Best Local Similarity   99.8%;
Matches  999;  Conservative   0;  Mismatches  1;  Indels  1;  Gaps  1;

Qy      1  tcgcagccacaagatccaaaccaagctccaacctcatcccacccggacatctcccccgct  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200  TCGCAGCCACAAGATTCAAACCCAGCTCCAACCTCATCCCACCCGGACATCTCCCCCGCT  1259

Qy     61  caccagttccgctgtaagaggaggagggccccacccccagtgtgccaatgttcagggtga  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260  CACCAGTTCCGCTGTAAGAGGAGGAGGGCCCCACCCCCAGTGTGCCAATGTTCAGGGTGA  1319

Qy    121  atgcctccctcaagtaccagctccgtcccagagaggagggtgcagagggatgccattatac  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320  ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGAGGAGGGTGGCAGAGGGATGCCATTATTAC  1379

Qy    181  ttgcaatcctgaggaattcatagttcagagtcgcgtgcagcttcccaaacttccagcgagcgt  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380  TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAAACTTCCAGCAGGCGT  1439

Qy    241  gcaggagtacaggagagtgcgcagacgagcgcccagcgccagcagagagaagaagtcagta  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440  GCAGGAGTACAGGAGAGTGCGCAGACGGCGCCCAGCGCCAGCAGAGAGAAGAAGTCAGTA  1499

Qy    301  cccagaaatcatcttcccttgaacaggttctgccataccccggatgaaagaaaatgtcag  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500  CCCAGAAATCATCTTCCTTGGAACAGGTTCTGCCATACCCCGGATGAAAGAAAATGTCAG  1559

Qy    361  tgccacacttgtcaacataagccccgacacgtctctggctactggactgtgtgagggcac  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560  TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGGCTACTGGACTGTGTGAGGGCAC  1619

Qy    421  atttgggcagctgtgccgtcattacggagaccaggtggacagaggtcctgggcaccctggc  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 3
BC004158
LOCUS   BC004158   3006 bp   mRNA   PRI   16-MAR-2001
DEFINITION Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:2441, mRNA, complete cds.
ACCESSION   BC004158
VERSION   BC004158.1 GI:13278770
KEYWORDS   MGC.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3006)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT   NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amdan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
REMARK   Clone distribution: MGC clone distribution information can be found

            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: c Column: 13.

FEATURES   Location/Qualifiers
     source   1..3006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:2441 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="vector: pOTB7"
     CDS   70..2550
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH04158.1"
                     /db_xref="GI:13278771"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVLQVVAAGSRDSGAALVFSEPNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYOIPIHSEORGKHOPWOSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRGVRDSSLVVAFICKLHLHKRGNFLVLKAKEMGLPVGPRAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDGAAFVVVECPDESFIOPICENATFORYQGKRAD
                     APVALVVHMAPASVLVDSRIYOQWMEREFGPDTQHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIPFLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPREWORDAIITCNPEEFIVEA
                     ISSLLRTCDLEEFOTCIVRHCKHAFGCALVEHSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKYPLFSPNFS
                     EKVGVAFDHHKVCFGDFPTMPKLIPLKALFPAGDIEEMEREREKRELRQVRAALLSRE
                     LAGGLEDGEPQOKRAHTEEPOAKKVRAQ"

BASE COUNT   742 a   808 c   858 g   598 t
ORIGIN

Query Match   98.6%;   Score 987.4;   DB 91;   Length 3006;
Best Local Similarity 99.8%;   Pred. No. 2.8e-261;
Matches 999;   Conservative   0;   Mismatches   1;   Indels   1;   Gaps   1;

Qy   1   tcgcagccacaagattcaaacccagtctcaaactcatccaccggagacatcttcccctgct   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209   TCGCAGCCACAAGATTCAAACCCAGTCTCAAACTCATCCACCGGAGACATCTTCCCCTGCT   1268

Qy   61   caccagcttccgctgtaagacggaggagggccccaacctcagtgtgccatggttcagggtga   120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269   CACCAGTTTCCGCTGTAAGAGACGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGA   1328

Qy   121   atgcctcctcaagtaccagctcccgtcccagaggaggggatggcagaggggatgccattac   180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329   ATGCCTCCTCAAGTACCAGCTCCCGTCCCAGAGGAGGGGATGGCAGAGGGGATGCCATTATTAC   1388

Qy   181   ttgcaatcctgaggagttcatagttgaggcgctgcagcttccacttccagagagcgt   240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389   TTGCAATCCTGAGGAGTTCATAGTTGAGGCGCTGCAGCTTCCACTTCCAGAGAGCGT   1448

Qy   241   gcaggagtacaggaggagtgcgcaggacggcccagcccagcagagaagaaagaagtcagta   300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449   GCAGGAGTACAGGAGGAGTGCCAGGACGGCCCAGCCCAGCAGAGAGAAAGAAGTCAGTA   1508

Qy   301   ccagaaatcatcattccttggaacaggggtctgccatcccggtgaagattcgaaatgtcag   360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509   CCCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATCAAGAGATTCGAAATGTCAG   1568

Qy   361   tgccacaacttgtcaacataagcccgacacgtctctgctactggactgtgtgagggcac   420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1569   TGCCACAACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGTGAGGGCAC   1628

Qy   421   atttgggcaagctgtgccgtcattacggagacaggttcctggggcaccctggc   480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1629   GTTTGGGCCAGCTGTGCCGTCATTACGGAGACCAGGTTCCTGGGGCACCCTGGC   1688

Qy   481   tgctgtgtttgtgtccccacctggcacgcagatcaccacacgggcttggcacgtactcttgc   540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 4
LOCUS   AF308698   2908 bp   mRNA   PRI   27-FEB-2001
DEFINITION Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION   AF308698
VERSION   AF308698.1 GI:10946496
KEYWORDS   .
SOURCE   chimpanzee.
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2 (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   Pan troglodytes ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES   Location/Qualifiers
     source   1..2908
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     CDS   1..2481
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24920.1"
                     /db_xref="GI:10946497"
                     /translation="MWALCSLLRSAAGRPMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGGCSGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGIQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEVEDETMYVYQIFIHSEQRRGHQPWQSPERPLSRLSPRSSSDSESSEN
                     ENEHLPHGVSQRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIRAPIIA
                     AVKDGKSITHEGRRILAEELCTPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHNAPESVLVDSRYQQWMERFGPDTQHLVLNENCASVNLRSHKIQTNLNI
                     HPDIFPLLTSFPCKKEGPFLSVPMVQGECLLRKTQYLPRREWQRDAIITQNPEEFIEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQDVRLGTLAAVFVSHLHADHHGLLNILLQRERAL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGNKVVYSGDIMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKIHSTTSQAISVGMRMRNAEFTMLNHFSQRYAKVPLFSPNFN
                     EKVGVAFDHMKVCFGDFATMPKLLIPPLKALFAGDIEEMERREKRELRQVRAALLSRE
                     LAGGLEDGEPQOKRAHTEEPQAKKVRAQ"
BASE COUNT   712 a   788 c   819 g   589 t
ORIGIN
Query Match   97.2%;   Score 973;   DB 89;   Length 2908;
Best Local Similarity 98.9%;   Pred. No. 2.6e-257;
Matches 990; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

Qy   1 tcgcagccacaagattcaaaccagctcaacctcatccaccacccggacaatcttccccgt 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 TCGCAGCCACAAGATTCAAACCAGCTCAACCTCATCCACCACCCGGACAATCTTCCCCCGCT 1199

Qy   61 caccagtttccgctgtaaagaaaggagggcccacccaacccagtgttgccatggttcagggtga 120
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   1200 CACCAGTTTCCCGCTGTAAGAAAGGAGGGGCCCACCCTCAGTGTTGCCCATGGTTCAGGGTGA 1259

Qy   121 atgcctcctcaaagtaccagctccgtccccaggagaaggagtgcagggaagatgccatt attac 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   1260 ATGCCTCCTCAAAGTACCAGCTCCGTCCCCAGGAGGAGGAGTGCCAGGGAGATGCCATTATTAC 1319

Qy   181 ttgcaatcctgagagaattcatagttgaggcgctgcagcttccaacttccagcagagcgt 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 TTGCAATCCTGAGAGAATTCATATAATTGAGGCGCTGCAGCTTCCAACTTCCAGCAGAGTGT 1379

Qy   241 gcaggagtacagggaggagtcgcgcaggaccgacggcccagcagagacagagaaagaaagtcaagta 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 GCAGGAGTACAGGGAGGAGTCGCGCAGGACGGCCCAGCAGAGAGACAGAGAAAGAAAGTCAGTA 1439

Qy   301 cccagaaatcatcctccttggaacacaggtgtcgcatccgatgaagattcgaaatgtcag 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 CCCAGAAATCATCTTCCTTGGAACACAGGTGTCGCATCCCGATGAAGATTCGAAATGTCAG 1499

Qy   361 tgccaacttgtcaacataagcccgacacagtctctgctactggactggtggaggcac 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500 TGCCACACTTGTCAACATAAGCCCCGACACAGTCTCTGCTACTGGACTGGTGGAGGCAC 1559

Qy   421 atttgggcagctgtgccgtcaattacggagacccaggtcctgggcacccctggc 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 GTTTGGGCAGCTGTGCCGTCATTACGGAGACCCAGGTCCTGGGCACCCCTGGC 1619

Qy   541 tgcaggagaacgcgccttggcatcttgggaaagccgcttcaccctttgtcggtgttg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1689 TGCTGTGTTTGTGTCCCACCT-CCACGCAGATCACCACACGGCGCTTGCCAAGTATCTGC 1747

Qy   601 cccccaaccagctcaaagcctggctccagcagtaccaacaacagtgccaggaggtcctgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1748 TGCAGAGAGAACGCGCGTTTGGCATCTTTGGGAAAGCCCGCTTCACCCTTTGCTGGTGTTG 1807

Qy   661 accacattcagtatgattcctgccaaatgcctcaggaagggggctgagatctccagtcctg 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1808 CCCCCAACCAGCTCAAAAGCCTGGCTCCAGCAGTACCAACAACAGTGCCAGGAGGTCCTGC 1867

Qy   721 cagtggaaagattgatcaagttcgcgtcttgtgcaacatgtgattttggagaagttttcagacct 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1868 ACCACATCCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTG 1927

Qy   781 gtctgtgtgcggacaccgcaagacgtcgtgctggtcgcaacacctctggctgga 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1928 CAGTGGAAAGATTGATCAGTTCGCGTTGTGCGCAACATGTGATTTTGGAGAAGTTTTCAGACCT 1987

Qy   841 aagtggttctattccggggaccacatcgccctcgccagagctctgtgtgtcgtgtgcacaccttggctggtga 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1988 GTCTGTGTGCGGCACTCCAAGACATGCGTTTGCGTTGTGTCGTCGCAACACCTCTGGCTGGA 2047

Qy   901 ccaccctcctgatacatgaagccacctgaagctgaagatgtgtttgaagaggaagcagtgaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2048 AAGTGGTTCTATTCCGGGGACCACATCGCCCTCGCCAGAGCTCTGTGTGTCCGGGATACATG 2107

Qy   961 agacacacagacaacgtcccaaagccatcagcagcgtgggatg 1001
          |||||||||||||||||||||||||||||||||||||||||||
Db   2108 CCACCCTCCTGATACATGAAGCCACCTGAAGCTGAAGATGTGTTTGAAGAGGAAGCAGTGAAA 2167

Db   2168 AGACACACAGACAACGTCCCAAGCCATCAGCGTGGGATG 2208

```
Qy    481 tgctgtgttgtggtcccacctggcacgcagatcaccacacgggggcttgccaagtatcttgc 540
          |||||||||||||||||||||||||||||||||||||| |||| ||||| ||||||||||||
Db   1620 TGCTGTGTTGTGTCCCACCT-GCACGCAGATCACCACACGGGGCTTGCTAAATAYCTTGC 1678

Qy    541 tgcagagagaacgcgccttggcatctcttgggaaagccgcttcacccttgctgtggttg 600
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   1679 TGCAGAGAGAACGACCCTTGCATCTTTTGGGAAAACCCTTTCACCCTTGCTGTGGTTG 1738

Qy    601 ccccaaaccagctcaaagcctggctccagacagtaccaacaacaagtggccagggaggtcctgc 660
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1739 CCCCCAACCAGCTCAAAGCCTGGCTCCAGACGTACCAACAACAAGTGGCCAGGGAGGTCCTGC 1798

Qy    661 accaacatcagtatgattcctgcccaaatgccttcaggaggggggctgagaatctcagtccctg 720
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   1799 ACCAACATCAGTATGATTYCTTGCCCAAAATGCCTTCAGGAGGGGGCTGAGAATCTCCAGTCCCTG 1858

Qy    721 cagtggaaagattgatcagttcgtgttggcgaacacatgtgatttggaagagttcagacct 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1859 CAGTGGAAAGATTGATCAGTTCGTCGTTGGCGAACAtGTGAtTTGGAAGAGTTTCAGACCT 1918

Qy    781 gtctgtgtgggcactgcagcaatgcgtttggcgtgtgtcgcgtggcacactctggctgga 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1919 GTCTGTGCGGGCACTGCAGCAAGCATGCGGTTGGCGTGTGTCGCGTGGCACACTCTGGCTGGA 1978

Qy    841 aagtgtgtctattccgggggaccacatagccctcgcgaggctctggtccggatgggggaaagatg 900
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1979 AAGTGGTCTATTCCGGGGACCACCATGCCCTCGCGAGGCTCTGGTCCGGATGGGGAAAGATG 2038

Qy    901 ccaccctcctgatacatgaagcaaccctgggaagatgtgttggaagggaagacagtggaa 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2039 CCACCCTCCTGATACATGAAGCAACCCTGGGAAGATGTGTTGGAAGGGAAGGGAAGACAGTGGAA 2098

Qy    961 agaacaacagcaacgtcccaagccaatcaagcgtggggatg 1001
          |||||||||||||||||||||||||||||||||||||||||
Db   2099 AGACACAACAGCAACGTCCCAAGCCATCAAGCGTGGGGATG 2139
```

```
RESULT  5
LOCUS       AK001392          2976 bp    mRNA          PRI        22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION   AK001392
VERSION     AK001392.1  GI:7022621
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2000985.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished (2000)
REFERENCE   2  (bases 1 to 2976)
AUTHORS     Isogai,T. and Otsuki,T.
TITLE       Direct Submission
JOURNAL     Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; cDNA library construction; clone selection; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center,
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
```

```
FEATURES             Location/Qualifiers
     source          1..2976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RP2000985"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pMEl8SFlj3; mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             75..2639
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91666.1"
                     /db_xref="GI:7022622"
                     /translation="MWALCSLLRSAAGRTMSQGRIISQAPARREPRPRKDPLRHLRTRE
                     KRGPSGCSGGPFNTVLQVVAAGSRDSGAALYVFSEPNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAMRPHSAPEYEDETMTVYQIPIHSEQRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRGVRDSSLVVAFICKLHLKRGNFLVLAKAKEMGLPVGTAAIAPTIA
                     AVKDKSITHGRELLAEELCTPPDGAAFVVVECPDESFTQPICENATFQRYQKRAD
                     APVALVVHMAPASVLLVDSRIYQQWMEREGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     QLPNFQQSVDQBYRRSAODGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISPD
                     TSLLDCGEGTFGQLCRHYGPDOVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERALA
                     SLGKPLHPLLVVAPNQLKAWLQQYRMNQOQEVLHIHSMIPAKCLQEGAEISSPAVERLI
                     SSLLRTCDLLEEFQTCLVRHCHHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDAPLL
                     IHEATLEDGLEEEAVEKTHSTTSOAISVGMRMNAEFIMLNHESORYAKVPLFSPNFSE
                     KVGVAPFDHMKVCLGDEPTMRKLIIPDLKALFAGDIEEMEERERELRQVRAALLSREL
                     AGGLEDGSLRSGPTQRSHRPRSEPSEDLGDPELRRLCYECPTHAPVSALLAGRS"
```

```
BASE COUNT      712 a      807 c      856 g      601 t
ORIGIN
```

```
Qy   Query Match     97.0%;  Score 971.4;  DB 89;  Length 2976;
     Best Local Similarity  99.5%;  Pred. No. 7.1e-257;
     Matches  996;  Conservative  0;  Mismatches  1;  Indels  4;  Gaps  2;

Qy      1 tcgcagccacaagattcaaaccccagctcaacctcatccacccggacaatctttccccctgct 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1214 TCGCAGCCACAAGAATTCAAAACCCCAGTCAACACCTCATCCACCCGGACAATCTTCCCCCTGCT 1273

Qy     61 caccagcgtttccgctgtaagaaggagagggcccccacccctcaagtgtgccccatggttcagggtga 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1274 CACCAGTTTCCGCTGT---AAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGA 1330

Qy    121 atgcctcccttcaagtaccagctcccgtcccaggagggggtggcagaggagggatgcattac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 ATGCCTCCTTCAAGTACCAGCTCCGTCCCAGGAGGGGAGTGGCACAGGGATGCCATTATTAC 1390

Qy    181 ttgcaatcctgagggaattcatcagttgaggcgcctgcagctcgagctcccaacttccagcagagcgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 TTGCAATCCTGAGGGAATTCATNAGTTGAGGCGCCTGCAGCTTCCCAACTTCCACCAGAGCGT 1450

Qy    241 gcaggaagtacaggaggagtgcgcaggagcggccagcaagcagaagaagaaagaagtcagta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 GCAGGAGTACAGGAGTGCCACGGACGGCCCAGCCAGCAAGAGAGAAAGAAGTCAGTA 1510

Qy    301 cccagaaaatcatcttcttggaacagggttctgccaatccccgatgaagattcgaaatgtcag 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCCAGAAAATCATCTTCCTTGGAACAGGGTTCTGCCAATCCCGATGAAGATTCGAAATGTCAG 1570

Qy    361 tgccacacttgtcaacataagccccgacacgtctctgctactgtggtgagggcac 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACGTGGTGAGGGCAC 1630

Qy    421 atttggcagctgtccgtcattacgggagacaggagcaggtggacagggtctgggccaccctggc 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1631 ATTTGGCAGCTGTGCCGTCATTACGGGACACAGGTGGACAGGGTCTGGGCCACCCTGGC 1690

Qy    481 tgctgtgtttgtgtcccacctggcacgcagatcaccacacgggggcttgccaagtatcttgc 540
```

RESULT 6

```
LOCUS       AF308694                2893 bp    mRNA            PRI       27-FEB-2001
DEFINITION  Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION   AF308694
VERSION     AF308694.1  GI:10946488
KEYWORDS    .
SOURCE      gorilla.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
```

```
REFERENCE   1  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
REFERENCE   3  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Gorilla gorilla ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
```

```
FEATURES             Location/Qualifiers
     source          1..2893
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
     gene            1..2893
                     /gene="ELAC2"
     CDS             1..2481
                     /gene="ELAC2"
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24916.1"
                     /db_xref="GI:10946489"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCCSGDPNTVVLQVVAAGSRDSGAALYVFSEENRYLFNCGEGVQRLMQEHKLKV
                     VRLDNIFLTRMHWHSNVVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRGRHGHPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTVAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAPVVVECPDESFIOPICENATFORYOGKAD
                     APVALLVHMAPESVLVDSRYQOWMRFGPTQOHLVINENCASVHNIRSHKIQTOLNLI
                     HPDIFPLLTSFPCKKEGPTLSVPMVQGECLLKYQLRPRREWORDAIITCNPEEFIVEA
                     LQLNFOQSVOEYRRSVODVPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     ASLGKPLHPLLVVAPSQLKAWLQQYVHNQCQEVLIHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFOTCLVRHCKHAFGCALVHTSGNKVVVSGDTMPCEALVRMGKDATL
                     LIHRATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKAVPLFSPNEN
                     EKVGVAFPDHNKVCFGDFPTMPKLIPPLKALFAGDIEEMERREKELRQVRAALLSGE
                     LAGGLEDGEPQOKRAHTEEFQOAKKVRAQ"
```

```
BASE COUNT      705 a    787 c    815 g    586 t
ORIGIN
```

```
Query Match    96.4%;  Score 965;  DB 89;  Length 2893;
Best Local Similarity  98.4%;  Pred. No. 4.1e-255;
Matches 985;  Conservative  0;  Mismatches  15;  Indels  1;  Gaps  1;

Qy        1 tcgcagccacaagattcaaacccagcgctcaacctcatccaccggacatcttcccccctgct 60
            |||||||||||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db     1140 TCGCACCACAAGATTCAAACCCAGCTCAACCTCATCCACCGGACATCTTCCCCCTGCT 1199

Qy       61 caccagtttccgctgtaagagaggagggccccaccccccagtgtgccatggttcaggtga 120
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     1200 CACCAGTTTCCCCTGTAAGAAGGAGGAGGGCCCCACCCCTCAGTGTGCCCATGGTTCAGGGTGA 1259

Qy      121 atgcctcctcaagtaccagctcccgtcccagaggagggggtggcagaggatgccattattac 180
            ||||||||||||||||||||||||||||||||| |||| ||||||||| |||||| |||
Db     1260 ATGCCTCCTCAAGTACCACGCTCCGTCCCAGAGGAGGAATGGCAGAGGGATGCCATTATCAC 1319

Qy      181 ttgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacctccagcgagcgt 240
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||| ||
Db     1320 TTGCAATCCTGAGGAATTCATAGTTGAGGCCGCTGCAGCTTCCCAACTTCCAGCAGAGTGT 1379

Qy      241 gcaggagtacagaggagtgcgcaggacggccaccagcaagaggaaagaagtcagta 300
            |||||||||||||| ||||||||| || ||||||| |||| ||||||||||||||
Db     1380 GCAGGAGTACAGAGGAGTGTGCAGGACAGTCCCAGCAGCAGGAAAAGAGTCAGTA 1439

Qy      301 cccagaaatcatcttcccttgaacaggtctgccatcccgatgaagaagatcgaaatgtcag 360
            ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     1440 CCCAGAAATCATCTTCCTTGAACAGGTCTGCCATCCCGATGAAGAAGATCCAAAATGTCAG 1499

Qy      361 tgccaacttgtcaacataagcccgacacgtcttgctactggactgtggtgagggcac 420
```

```
Db  1500 TGCCACACTTGTCAACATAAGCCCGGACAGGTCCTGCTACTGGACTGTGGTGAGGGCAC 1559
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   421 atttgggcagctggccgtcattacggagaccaggtggacaggtcctggcaccctggc 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1560 GTTTGGGCAGCTGTGCCGTCATTACGGACAGGGGTTCTGGGCACCCTGGC 1619

Qy   481 tgctgtgttgtgttcccacctggcacgcaggatcacacagggggcttgccaagtatcttgc 540
         ||||||| |||||||||||||||||||||||||||||||| |||||| |||||||||
Db  1620 TGCTGTGTTGTGTGTCCCACCT-GCACGCAGATCACCACACGGGGCTTGCTAAATATCTTGC 1678

Qy   541 tgcagagagaacgcgcctggcatctttggggaaagccgcttcaccctttgctgtggttg 600
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1679 TGCAGAGAGAACAAGCCTTGGCATCTTTGGGAAAGCCCCTTCACCCTTTGCTGGTGGTTG 1738

Qy   601 cccccaaccagctcaaagcctggctccaggacgtaccaacaccaagtgccaggaggtcctgc 660
         |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1739 CCCCCAGCCAGCTCAAAAGCCTGGCTCCAGGACGTACCAACACCAAGTGCCCAGGAGGTCCTGC 1798

Qy   661 accaacatcagtatgattcctgccaaaatgctcaggaaggggctgagaatccagtcctg 720
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1799 ACCAACATCAGTATGATTCCTGCCAAAATGCTTCAGGAAGGGGCTGAGAATCCAGTCCTG 1858

Qy   721 cagtggaagagattgatcagttcgctgttgcgaacatgtgattggagagttcagacct 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1859 CAGTGGAAGAGATTGATCAGTTCGCTGTTGCGGACACATGTGATTGGAGAGTTTCAGACCT 1918

Qy   781 gtctggtgcgcactgcaagcaatggcgtttggctgtctggtgcacacctctggctgga 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1919 GTCTGGTGCGCGACTGCAAGCAATGGCCGTTTGGCTGTGTGGTGCACACCTCTGGCTGGA 1978

Qy   841 aagtggttctattccgggggacaaccatgccctgcgaggctctgttcggatggggaaatg 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1979 AAGTGGTTCTATTCCGGGGGACAACCATGCCCTGCGAGGCTCCGCACCTGGGGAAAGATG 2038

Qy   901 ccaccctcctgatacatgaagccaccctggaagatgtggtttgggaggagaagcagtggaaa 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2039 CCACCCTCCTGATACATGAAGCCACCCTGGAAGAGATGGTTTGGAACGAGAAGCAGTGGAAA 2098

Qy   961 agacacaagcacacaacgtcccaagccaatcaagcgtggggatg 1001
         ||||||||||||||||||||||||||||||||||||||||||||
Db  2099 AGACACAAGCACACAACGTCCCAAGCCCATCAAGCGTGGGGATG 2139
```

```
Qy     2 cgcagccacaagagattcaaacccagcctcaaccttcaccccgggacatcttccccctgctc 61
         ||||||||||||||||||||||||||||||||||||||||||||||| |||||| |||
Db  1143 CGCAGCCGCACAAGAGATTCAGACCCGGTCAGCCTCATCCACCCTGACATCTTCCCCCAGCTT 1202
```

Query Match   72.8%; Score 728.8; DB 94; Length 2712;
Best Local Similarity   83.7%;
Matches 837; Conservative   0; Mismatches 162; Indels   1; Gaps   1;

RESULT 8

LOCUS   AC005277   118788 bp   DNA   PRI   23-JUL-1998
DEFINITION   Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
ACCESSION   AC005277
VERSION   AC005277.1 GI:3337311
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 17, clone hRPK.597_M_12
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
             Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
             Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
             Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
             FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
             Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
             Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
             Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
             Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
             Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
             Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
             Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
             Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
             Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
             Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
             Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
             FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
             Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
             Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
             Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
             Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
             Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
             Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
             Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
             Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jul 23, 1998 this sequence version replaced gi:3335015.
             All repeats were identified using RepeatMasker: Smit, A.F.A. &
             Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             Only the first 118.8 kilobases of this clone are being submitted.
             The remainder overlaps accession number AC005274 (WICGR project
             L350).
FEATURES   Location/Qualifiers
     source   1..118788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.597_M_12"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   795..884
                     /rpt_family="MIR"
     repeat_region   complement(932..1000)
                     /rpt_family="MIR"
     repeat_region   2044..2076
                     /rpt_family="AT_rich"
     repeat_region   2237..2713
                     /rpt_family="MER34"
     repeat_region   2718..3477
                     /rpt_family="L2"
     repeat_region   complement(3478..3821)
                     /rpt_family="MLT1A1"

Qy   62 accaagttccgcgtgtaagaaggaaggggccccaacctcagtgtgcccatggttcaggtgaa 121
              |||||||||| |||
Db   1203 ACCAGCTTCTATAGTAAGGAGGAAGGGGTTCCACCCTCAGCGTGCCAACAGTTCGGGTGAA 1262

Qy   122 tgcctcctcaagtaccacgctccgtccaggaggggagtggcagagaggaggatccatttact 181
              |||||||||||||||||| ||| |||
Db   1263 TGCCTCCTCAAGTACTCAGCTCCGCCCAAAGAGAGAGAGTGGCAGAGGGATACCACCCGAC 1322

Qy   182 tgcaatcctgagggaattcattaagttgaggcgctgcaagcttccaacttccagcagagcgtg 241
              ||||| |||| ||| ||||||| ||||| ||
Db   1323 TGCAATACTGATGAATTCATTAACGTTAGCTTTGGACTTCCAGGAGGTGTG 1382

Qy   242 caggagtacaggagagggagtgccgcaggacgggcccagcccagcagagaaagaagtcagtac 301
              ||| |||| |||| ||||
Db   1383 GAGGAGTATCGGAAGAACGTGCAGGAAAACCAGCCCCAGCCAGTTCCAGGAGTCTG 1442

Qy   302 ccagaaatcatcttcctggaacaggggtctgccatccgataagaatcgaaatgtcagt 361
              | |||| ||| ||||||||
Db   1443 CCTGAAAATTGTCTTCCTGGGTACGGGGTCTGCCATCCCAATGAAAGATCCGAAATGTCAGT 1502

Qy   362 gccacaacttgtcaacataagcccgacaacgtctctgctctactggactgtgggggcaca 421
              |||||| |||| || || ||||
Db   1503 TCCACACTCGTCAACTTAAGCCCTGACAAGTCAGTGCTTCCTGGATTGTTGGAGAAGGCACT 1562

Qy   422 tttggtccagctgtgccgtcatcatacggaagacaggtgtgacaaggtcctgggcaccctggct 481
              |||||||| ||||||| | |||
Db   1563 TTTTGGGCAGTTGGTGCCGTCATTTACGGAGCACAGCAAATAGACCGAGTCTTTATGCAGCCCTCACG 1622

Qy   482 gctgtgttttgtcccaccacggcacgcagatccacaacgggcttgccaagtatcttgct 541
              ||||||||||||| || || |||
Db   1623 GCTGTGTTTGTGTCCCACCT-GCACGCCGACATCCACACGGGCCTTGCTGAATATCTTGCT 1681

Qy   542 gcagaagagaaccgccttggcatcttgggaaagccgcttcaccccttgctggtggttgc 601
              |||||| |||||| || || ||
Db   1683 GCAGAGAGAGCATGCCGTTGGCATCTCTGGGAAACCCTTGCTTGTGGTGGC 1741

Qy   602 cccaccagctcaaagccctggctccagcagtaccaccaccagtgccaaggaggtcctgca 661
              |||||||||||||| || ||
Db   1683 GCAGAGAGAGCATGCCGTTGGCATCTCTGGGAAACCCTTGCTTGTGGTGGC 1741

Qy   662 ccacatcagtatgattcctgcccaaatgcctttcaggaaagggagggggctgagatctccagtgcctcctgc 721
              ||||||||||||||| || |||
Db   1742 TCCTACCCAGCTCAGGGCCTGCACCAGTATTCACAACCACCTGCCACAGAGAATTCTGCA 1801

Qy   722 agtggaaagattgatacagttcgctgttgcggaacaatgttcggaaggagggttttcaagaccctg 781
              |||||||||| || |||| ||||||| ||||
Db   1803 CCACGTCAGTATGAATTCCTGCGCAAATGCCTTCAGAAAGGGGCAGAGGGTGTCCAATACTAC 1861

Qy   782 tctggttgcggcaactgcaagcatgcgtttggctctgtgccgttgcgcacacctctggaa 841
              |||||| ||||||||| ||||| |||
Db   1862 ATTGGAAAGGCTGATAAGCTTCTCGTGTTGGAAACATGTGACTTAGAAGAGTTTCAGACCTG 1921

Qy   842 agtggtctattccgggacacaacaatgccctgcaggcctcgggatggggagaaagatgc 901
              ||||||| || ||||||||||| ||
Db   1922 CCTGGTACGGCACTGCAAGCACTGCTTTTTGGCTGCTGCACTGGTCATTCATCTGGCTGGA 1981

Qy   902 caccctcctgatacatgaaccaccctggaaagatggtttggaagaggaagcagtggaaa 961
              ||||||||| || || |||||
Db   1982 AGTCGTCTACTCCGGGGATACCATGCCCTGGTCTGGTGCTCCTCGGATATGGGGAAAGATGC 2041

Qy   962 gacacacagcacaaacgtcccaaagccatcaagcgtgtggggatg 1001
              ||||||||| |||||
Db   2042 CACCCTCCTGATACATGAAGCCACTCTGGAGGATGGCTTGGAGGAGCAGTAGAGAA 2101

Db   2102 GACACACAGCACAACCTCCCAGGCCAATTAAATGTGGGGATG 2141

```
repeat_region   3822..4007                   /rpt_family="L2"
repeat_region   complement(4015..4281)       /rpt_family="MER89"
repeat_region   complement(4282..4321)       /rpt_family="(TAAA)n"
repeat_region   complement(4328..4608)       /rpt_family="AluSx"
repeat_region   complement(4609..4895)       /rpt_family="MER89"
repeat_region   complement(4896..5186)       /rpt_family="AluSp"
repeat_region   5219..5479                   /rpt_family="L2"
repeat_region   complement(5481..5740)       /rpt_family="AluYb8"
repeat_region   complement(5963..6275)       /rpt_family="MER34"
repeat_region   complement(6411..6494)       /rpt_family="AluSp"
repeat_region   6830..7013                   /rpt_family="MER5A"
repeat_region   7014..7318                   /rpt_family="MLT1C"
repeat_region   7319..7581                   /rpt_family="AluYb8"
repeat_region   7721..8032                   /rpt_family="MLT1C"
repeat_region   8039..8333                   /rpt_family="AluJb"
repeat_region   9431..9460                   /rpt_family="AluSq"
repeat_region   9662..9829                   /rpt_family="AT_rich"
repeat_region   complement(11302..11728)     /rpt_family="MIR"
repeat_region   complement(11824..11884)     /rpt_family="MER39"
repeat_region   complement(11928..12108)     /rpt_family="LTR16C"
repeat_region   12453..12725                 /rpt_family="LTR16C"
repeat_region   13377..13785                 /rpt_family="AluJb"
repeat_region   14118..14153                 /rpt_family="MSTB"
repeat_region   complement(15320..15620)     /rpt_family="(CA)n"
repeat_region   15683..15759                 /rpt_family="AluJo"
repeat_region   complement(15838..15893)     /rpt_family="(CA)n"
repeat_region   complement(15915..16080)     /rpt_family="MER91A"
repeat_region   complement(16474..16544)     /rpt_family="L2"
repeat_region   17907..17920                 /rpt_family="L2"
repeat_region   18271..18311                 /rpt_family="L1ME"
repeat_region   complement(18368..18529)     /rpt_family="AluJb"
repeat_region   complement(18530..18560)     /rpt_family="(CA)n"
repeat_region   complement(18561..18681)     /rpt_family="AluJb"
repeat_region   18760..19065                 /rpt_family="AluSg"
repeat_region   complement(19638..19896)     /rpt_family="MIR"
repeat_region   complement(19971..20260)     /rpt_family="L1ME3A"
repeat_region   20263..20552

repeat_region   complement(20553..20816)     /rpt_family="MER58B"
repeat_region   complement(20821..21069)     /rpt_family="L1ME3A"
repeat_region   complement(21071..21100)     /rpt_family="L1M4"
repeat_region   complement(21101..21338)     /rpt_family="(TAAA)n"
repeat_region   complement(21348..21598)     /rpt_family="AluSx"
repeat_region   21799..22104                 /rpt_family="L1M4"
repeat_region   22151..22562                 /rpt_family="AluJo"
repeat_region   23978..24287                 /rpt_family="MSTA"
repeat_region   24852..25199                 /rpt_family="AluSq"
repeat_region   complement(25697..25885)     /rpt_family="L2"
unsure          25816..25895                 /rpt_family="MLT1D"
                                             /note="Single-stranded coverage."
repeat_region   complement(25994..26099)     /rpt_family="purine-rich"
repeat_region   complement(26122..26291)     /rpt_family="MLT1D"
repeat_region   26312..26637                 /rpt_family="L2"
repeat_region   26456..26480                 /rpt_family="(CAAAA)n"
repeat_region   complement(26576..26781)     /rpt_family="(GGGAA)n"
repeat_region   complement(26912..26990)     /rpt_family="MER6B"
repeat_region   26991..27290                 /rpt_family="MER21B"
repeat_region   complement(27291..27975)     /rpt_family="AluSx"
repeat_region   28034..28197                 /rpt_family="MER21B"
repeat_region   28266..28795                 /rpt_family="LTR28"
repeat_region   28843..28916                 /rpt_family="LTR28"
repeat_region   28920..29005                 /rpt_family="LTR28"
repeat_region   29140..29233                 /rpt_family="MIR"
```

```
Query Match   Best Local Similarity   13.7%;   Score 137;   DB 85;   Length 118788;
Matches 225;   Conservative 74.3%;   Pred. No. 5.3e-27;   Mismatches 0;   Indels 78;   Gaps 1;

Qy      668 cagtatgattcctgccaaatgccttcaggaagggctgagatctccagtcctgcagtgga    727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101444 CAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGCTGAGATCTCCAGTCCTGCAGTGGA 101503

Qy      728 aagattgatcagttcgctgtgtgcgaacatgtgattggaag------------------    768
            |||||||||||||||||||||||||||||||||||||||||
Db   101504 AAGATTGATCAGTTCGCTGCTGTTGCGAACATGTGATTTGGAAGAGGGTAAGGGCACAGCCGC 101563

Qy      769 ----------------------------------------------------------a    769
                                                                      |
Db   101564 AGGCATCATGGGGGCGAGGTGGGGAGCAGAGCTGCAGAGCCCTTCCAGCCCACCTTTCA 101623

Qy      770 gtttcagacctgtctggtgcggcactgcaagcatgcgttgggctgcgctggtgtgcacac   829
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101624 GTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTGGGCTGCGCTGGTGTGCACAC 101683

Qy      830 ctctggctggaaagtcggtctcattccgggggacaccacgacccctcgcgaggctctcgggggat   889
            |||||||||||||||||||||||||||||| ||||||||| |||||||||||||||||| |||
Db   101684 CTCTGGCTGGAAAGTCGGTCTCATTCCGGGGACACCACCATGCCCTCGCGAGGCTCTCGGTCCGGAT 101743
```

```
Qy    890 ggg 892
              |||
Db 101744 GGG 101746

RESULT   9
LOCUS       AF304371S2               740 bp    DNA             PRI       19-OCT-2000
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 17 and partial cds.
ACCESSION   AF304369 AF304371S2
VERSION     AF304369.1  GI:10880929
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     mRNA            <301..>439
                     /number=17
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     exon            301..439
                     /number=17
     CDS             <301..>439
                     /codon_start=2
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24440.1"
                     /db_xref="GI:10809931"
                     /translation="PDTSLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADH
                     HT"
BASE COUNT      154 a    188 c    173 g    225 t
ORIGIN

Query Match     13.5%;  Score 135.2;  DB 89;  Length 740;
Best Local Similarity   89.2%;
Matches 157;  Conservative 0;  Mismatches 18;  Indels 1;  Gaps 1;

Qy    347 attcgaaatgtcagtgccacacttgtcaacataagcccggacacgtctctgctactggac 406
          ||||  ||||| |||||||||||| ||||||||||||||||||||||||||||||||||||
```

```
Db  266 AGTGGCAGTGAGACTCTCTTCTTCTTCTTCTTCTTCTGCGAGCCCGACACGCTCTCTCCTGCTACTGGAC 325
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  407 tgtggtgagggcacacattgggcagctgtgccgtcgcattacggagaccaggtggacagggtc 466
Db  326 TGTGGTGAGGGCACACATTTGGGCAGCTGTGCCGTCGCATTACGGAGACCAGGTGGACAGGGTC 385
Qy  467 ctgggcaaccctggctgctgtgttgttctgtcccacctgggcaggcaggatcaccaacacgg 522
        |||||||||||||||||||||||||||||||||||||| ||||||||| ||||||| |||||||
Db  386 CTGGGCAACCCTGGCTGCTGTGTTGTTCTGTCCCACCT-GCACGGCAGATCACCAGATCACCACCACACGG 440

RESULT   10
LOCUS       AF348157               34593 bp    DNA             ROD       04-APR-2001
DEFINITION  Mus musculus putative prostate cancer susceptibility protein
            (Elac2) gene, complete cds, alternatively spliced.
ACCESSION   AF348157
VERSION     AF348157.1  GI:13540341
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Mouse Elac2-containing genomic DNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 34593)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..34593
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     mRNA            join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24652,
                     27291..27429,27763..27801,27930..28039,28553..28652,
                     28733..28853,29101..29179,29710..>30639)
                     /gene="Elac2"
                     /note="alternatively spliced 24 exon form"
                     /product="putative prostate cancer susceptibility protein"
     mRNA            join(<8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24652,
                     27291..27429,27763..27801,27930..28039,28553..28652,
                     28733..28853,29101..29179,29710..>31277)
                     /gene="Elac2"
                     /note="alternatively spliced 25 exon form"
                     /product="putative prostate cancer susceptibility protein"
     gene            <8470..>31277
                     /gene="Elac2"
```

CDS   join(8470..8690,9077..9127,9266..9336,9948..10012,
                10957..11014,12096..12164,12933..13064,14902..14960,
                16609..16667,16746..16818,17370..17482,18671..18766,
                21539..21677,21757..21842,23553..23671,24619..24715,
                27291..27429,27763..27801,27930..28039,28553..28652,
                28733..28853,29101..29179,29710..29854,30406..30639)
                /gene="Elac2"
                /note="24 exon form"
                /codon_start=1
                /product="putative prostate cancer susceptibility protein"
                /protein_id="AAK29421.1"
                /db_xref="GI:13540343"
                /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTREKRGPGP
                GGPNTVYLQVVAAGRDAGAALYFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLERAIKIFSGPLKGIELAVRPH
                SAPEYKDETMTVYQVPIHSERRCGKQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                GKSITYEGRETAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                LVHHAPESVLIDSRYQQWMERFGPDTQHILINENCPSVHNLRSHKIQTQLSLIHPDI
                FPQLISFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTLDCNTDEFIEAEALELP
                SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                LLDGCEGTFGQOLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASIG
                KPFQPLLVVAPTQLRAWLQOYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISL
                ATLEDGLEEBAAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                EDREPQOKRAHTDEPHSPQSKKVRAQ"

COMMENT   Details of yeast sequencing at the Sanger Centre are available on
                the World Wide Web.
                (URL, http://www.sanger.ac.uk/Projects/S_pombe)
                Protein coding regions (CDS) have been predicted with the help of
                computer analysis using the Genefinder program in PomBase (an ACEDB
                database) with additional predictions for the branch-acceptor sites
                supplied by the program Sp3splice. CAUTION: It is possible that for
                any individual CDS we may have underestimated or overestimated the
                number of introns/exons or we may not have chosen the correct
                splice donor/acceptor sites.
                CDS are numbered using the following system eg SPAC5H10.01c, SP (S.
                pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
                (complementary strand).
                The more significant matches with motifs in the PROSITE database
                are also included but some of these may be fortuitous.
                The length in codons is given for each CDS.
                IMPORTANT: This sequence MAY NOT be the entire insert of the
                sequenced clone. It may be shorter because we only sequence
                overlapping sections once, or longer, because we arrange for a
                small overlap between neighbouring submissions.
                Cosmid c1D4 is overlapped at the 5' end by cosmid c1F3 and at the
                3' end by cosmid c22F3.

FEATURES   Location/Qualifiers
    source   1..31201
                /organism="Schizosaccharomyces pombe"
                /strain="972h-"
                /db_xref="taxon:4896"
                /chromosome="I"
                /clone="cosmid c1D4"
                /map="L"

    misc_feature 1..133
                /note="the sequence of this region is derived from cosmid
                c1F3, the true end of c1D4 is at position 134, while the
                true end of c1F3 is at position 4137 in this sequence"

    gene   7..864
                /gene="SPAC1D4.01"
                /note="SPAC1D4.01"

    CDS   7..864
                /gene="SPAC1D4.01"
                /codon_start=1
                /note="hypothetical 33.0 kd protein"
                /product="hypothetical 33.0 kd protein"
                /protein_id="CAB56125.1"
                /db_xref="GI:5912524"
                /translation="MRQIFGTCTGYKMGIFQEGLLEIPTTKSIRNSIHIKKKSNRS
                FRRKVPGNEKEFDLEELDDNDIRLRQALEATKRKIRNSIIGINAEKLLNQETKKEK
                OLNTANEPHERANDQTSAQSSKLIEAQLPTVEDRFAKQTNEVDINTHLLNFVEKKLKOE"

CDS   join(8470..8690,9077..9127,9266..9336,9948..10012,
                10957..11014,12096..12164,12933..13064,14902..14960,
                16609..16667,16746..16818,17370..17482,18671..18766,
                21539..21677,21757..21842,23553..23671,24619..24715,
                27291..27429,27763..27801,27930..28039,28553..28652,
                28733..28853,29101..29179,29710..29854,30406..30624,
                31242..31277)
                /gene="Elac2"
                /note="25 exon form"
                /codon_start=1
                /product="putative prostate cancer susceptibility protein"
                /protein_id="AAK29420.1"
                /db_xref="GI:13540342"
                /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTREKRGPGP
                GGPNTVYLQVVAAGRDAGAALYFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLERAIKIFSGPLKGIELAVRPH
                SAPEYKDETMTVYQVPIHSERRCGKQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                GKSITYEGRETAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                LVHHAPESVLIDSRYQQWMERFGPDTQHILINENCPSVHNLRSHKIQTQLSLIHPDI
                FPQLISFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTLDCNTDEFIEAEALELP
                SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                LLDGCEGTFGQOLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASIG
                KPFQPLLVVAPTQLRAWLQOYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISL
                ATLEDGLEEBAAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                EDREPQOKRAHTDEPHSPQSKKESVANTLGARV"

Db   28851 TGGGTAA 28857

RESULT 11

LOCUS   SPAC1D4   31201 bp   DNA   PLN   15-OCT-1999
DEFINITION S.pombe chromosome I cosmid c1D4.
ACCESSION   Z69239
VERSION   Z69239.1 GI:1177333
KEYWORDS   SPAC1D4.01; byr1; cct2; CDP-alcohol phosphatidyltransferase; chaperonin;
            conjugation and sporulation; cskl; cyclin suppressing protein
            kinase; DNA repair helicase; golgi peripheral membrane protein
            protein kinase; rad15; rhp3; serine threonine protein kinase; stel;
            TCP1 beta subunit homologue; tf2 LTR.
SOURCE   fission yeast.
  ORGANISM Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1 (bases 1 to 31201)
  AUTHORS   Lye,G., Churcher,C.M., Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-FEB-1995) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
            Notes:

BASE COUNT   8641 a   8030 c   8142 g   9780 t
ORIGIN

Qy   769 agttcaagacctgctggtgcggacactggcactggcctgctggtgcaca   828
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   28731 AGTTCAGAGACCTGCCTGGTACGGCACTGCAAGCATGCTTTTGGCTGTGCACTGGTACATT 28790

Qy   829 ccctctggctggaaagtggtctattccgggacacacatgcccctgcaggcttggtccgga   888
             |||||||||||||||||||| |||||| |||||||||||||||||| |||||| ||
Db   28791 CATCTGGCTGGAGAAAGTCGTCTACTCGGGGGATACCAATGCCCTGTGAGGCCTCTGGTCCAGA 28850

Qy   889 tggggaa   895
             ||||| ||
Db   28851 TGGGTAA 28857

Query Match   88.2%; Score 103; DB 94; Length 34593;
Best Local Similarity   88.2%; Pred. No. 1.4e-17;
Matches 112; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
                    RLAQNYSENGETNALNTKNESTVQNIKNSLHPNEHSFIRDAAALGAIREVDLGIISTD
                    VDNLKNGRKRQKKRARMEKLDSKALPTSEDAARDEFIEKMLKPISQDDEESKGIYRF
                    RVKDGTQD"
gene            complement(join(1088..1636,1686..1873,1970..2090,
                2137..2316))
                /gene="SPAC1D4.02c"

CDS             complement(join(1088..1636,1686..1873,1970..2090,
                2137..2316))
                /gene="SPAC1D4.02c"
                /note="SPAC1D4.02c, len:345, SIMILARITY:Rattus norvegicus,
                O35254, golgi peripheral membrane protein p65., (451 aa),
                fasta scores: opt: 493, E():4.7e-20, (34.4% identity in
                305 aa)"
                /codon_start=1
                /product="putative golgi peripheral membrane protein"
                /protein_id="CAB56126.1"
                /db_xref="GI:5912525"
                /translation="MFGGLKNFIKEKSEALAGIHRESDESCGFRVLKVENDSKAYNAR
                IESYYDFITAVNGILLNGDPSMFMALLRDSSPEVTLEVFSLRGIQITRKVNIKINSDEK
                IGMVLQWASIAPAVDAIHHILNVIDDSPVARASLVPYEDYIVGTPEGMMTGEKALSDL
                IESHLNRPLRLIYINYHRDSTRQVTIVPNRHWGGNGAIGCGVGHGVLHRLPAPLSGPP
                PQPGDIVFSNPMLGGPDHKVSQPSETENFLPTPEPPKIASANAGSSNEISIPHYQRHK
                KSHKGAIODSSIOSYLDEEEKLSRELDHKTKDASSTNDSQTTPLPPPPVAVANSTNDE
                SAPQNEELVKN"
                /codon_start=1

misc_feature    complement(1637..1651)
                /gene="SPAC1D4.02c"
                /label=SPAC1D4.02c
                /note="splice branch and acceptor sequence,
                ctaacaattaacag"

misc_feature    complement(1680..1685)
                /note="splice donor sequence, gtatgt"

misc_feature    complement(1874..1886)
                /note="splice branch and acceptor sequence, ctaacgactacag"

misc_feature    complement(1964..1969)
                /note="splice donor sequence, gtaagt"

misc_feature    complement(2091..2108)
                /note="splice branch and acceptor"
                /note="ctaaagtgttttcacag, splice branch and acceptor"

misc_feature    complement(2131..2136)
                /note="gtatga, splice donor sequence"

gene            complement(join(2574..2820,2907..4201))
                /gene="SPAC1D4.03c"

CDS             complement(join(2574..2820,2907..4201))
                /gene="SPAC1D4.03c"
                /note="SPAC1D4.03c, len:513, SIMILARITY:Fugu rubripes,
                Q9YGN1, Sand protein., (520 aa), fasta scores: opt: 609,
                E():7.2e-32, (28.5% identity in 407 aa)"
                /codon_start=1
                /label=SPAC1D4.03c
                /product="hypothetical protein"
                /protein_id="CAA93212.1"
                /db_xref="GI:1177336"
                /translation="MEPTSEHSSIKEEVENDNVHRSHESECGSLLLNPGNVLMAAPSV
                SEDDEVSRSTPELRSHVENVEQLLSDILHDNSSPLNVSTSVSSSSNNTPAVDEIIKLL
                SLISFDLAKQKRTYLIFSSSGKPVFSNIVDDSIEPSTVGALQAIISSFEVSKEELTSF
                STFSNVVIVLSKNPLYLGVGVSPSTTLSAAYLSELNLLYCQILTGVTAKAMQLTLNSR
                PNFDLRRLTIGSNEQFLKELCDQLNDYELVEPTLNAISPLPLRSSPDDLSQLLLRETPK
                SLLFTFIAIRGRLVCMVKAKLLLHANDLYLLFLSLFRYQSFNDSMEHHVPVCPPTLN
                PDAYIYISYFLCKDTVLIMGSESGVFEMQSVKCKVAQEIQDHGWLKKLIYCEEMD
                RTTPRNPGSPCISHYLFYSKKYSQFYTPGYSFSTPNFNRTLYAYASLHDQAFHKKN
                SFSINMTVHESLLLFTWSTPASFEDPHCIANATTSSQLLIANVNKILRWIRREENLFIQ
                TNLSF"
                /codon_start=1

misc_feature    complement(2821..2836)
                /note="splice branch and acceptor sequence,
                ctaactatgtttta g"

misc_feature    complement(2901..2906)
                /note="splice donor sequence, gtacgt"

gene            join(5400..5402,5439..5570,5618..7066)
                /gene="SPAC1D4.04"

CDS             join(5400..5402,5439..5570,5618..7066)
                /note="cct2"
                /gene="SPAC1D4.04"
                /note="SPAC1D4.04, len:527"
                /codon_start=1
```

Right column:

```
                /product="probable t-complex protein 1, beta subunit"
                /protein_id="CAA93213.1"
                /db_xref="GI:1177337"
                /translation="MSLNPHQIFNESGIQERGENARLSSEVGAIAVGDLIVKSTLGPKG
                MDKILQSNSSDIVVTNDGATILKSIALDNAAKVLVNISKVQDDEVGDGTTSVCVFA
                AELLRQAEIMVNAKIHPQVIIDGFRSLAIKTAIDALRASSIDNSSDPAKFRSDLENIAR
                TTLSSKILSQNKNHFAQLAVDAVLRLKGSTHNIDNIQIIKIGGKIDDSFLDEGFINK
                TIGVNCPKVMENANILIANTAMDTDKVKVFGARVRVDTTGKLAELERAEREKMKAAVE
                KIKSININCINRQLIYNWPEQLFADAGIMSIEHADFDGTERLSLVTGGEIASTFDHP
                ELVKLGHCHKIEEIIGEDKMIKFSGVEAGEACTIVLRGATHQLLDESERAIHDALAV
                LSQTVAESRVTLGGCGAEMLMAKAVEEAATHEPGKKAVAVSAFAKALSQLPTILADNA
                GFDSSELVAQLKAAHYDGNDTMGLDMDEGEIADMRAKGILEALKLKQAVVSSGSEGAQ
                LLRVDTILKAAPRPRERM"

misc_feature    5425..5438
                /gene="SPAC1D4.04"
                /note="splice branch and acceptor sequence,
                tactaacgatatag"

misc_feature    join(5529..5570,5618..7045)
                /gene="SPAC1D4.04"
                /note="Match to PF00118 cpn60_TCP1, TCP-1/cpn60 chaperonin
                family. Score 590.77"

misc_feature    5571..5576
                /gene="SPAC1D4.04"
                /note="splice donor sequence, gtaagg"

misc_feature    5604..5617
                /gene="SPAC1D4.04"
                /note="splice branch and acceptor sequence,
                ttaacgatttatag"

misc_feature    5741..5767
                /gene="SPAC1D4.04"
                /note="PS00995 Chaperonins TCP-1 signature 3"

gene            complement(join(7282..8366,8419..8497))
                /gene="SPAC1D4.05c"

CDS             complement(join(7282..8366,8419..8497))
                /gene="SPAC1D4.05c"
                /note="SPAC1D4.05c, len:387, LOW SIMILARITY:Mus musculus,
                Q92OU0, polytropic murine leukamia virus receptor Syg1.,
                (695 aa), fasta scores: opt: 261, E():1.5e-10, (25.1%
                identity in 398 aa) also shows low similarity to,
                Saccharomyces cerevisiae, ERD1, YEAST, ERD1 protein., (362
                aa), fasta scores: opt: 187, E(): 1e-05, (23.6% identity
                in 382 aa)"
                /note="SPAC1D4.05c, len:387, SIMILARITY:Mus musculus,
                polytropic murine leukamia virus receptor syg1., (695 aa),
                fasta scores: opt: 261, E():1.5e-10, (25.1% identity in
                398 aa)"
                /codon_start=1
                /label=SPAC1D4.05c
                /product="hypothetical protein"
                /protein_id="CAA93214.1"
                /db_xref="GI:1177338"
                /translation="MDLEVVEFPLHIKLALPRIGLLIVIGTWLWSVCYHLIYILNRY
                QPISPNPRGSINSKWYHLLQIPLSNRHTDLEENTERKANLVSPVDFHAGYCFAAILSI
                SWATGFILFLKKTQGDIGGLYSHPITPLLMVTTAFILVFPFPWRRSSQRGLRKSII
                RVELFQOADFRSPYKDFIVSELFTSYAKALGDFYIFGCVLQHISKFTLRPDLKDCGT
                FFVPLAMAYPFIVAIQCLHYGLSRKHTFRINLLSALKHATALPVIYLSAIIHAKQT
                KFTLTGHGYLFWLWLLSALLSSAYFLWDVFIDWRIFRFPHKSINIKRFPMFIYAIG
                CFINFILRVTWSMKLHPRLHQFHEYEMGIFSFEMLEILRFEIWLFFHLDAISS"
                /codon_start=1

misc_feature    complement(8367..8389)
                /note="splice branch and acceptor sequence,
                ttaacgatttatag"
```

Bottom alignment section:

```
Query Match     7.18;  Score 71.4;  DB 15;  Length 31201;
Best Local Similarity 54.9%;
Matches 141;  Conservative 0;  Mismatches 116;  Indels 0;  Gaps 0;

                       Score 71.4;  Pred. No. 7e-09;

Qy     740 ttcgctgttggaacatgtgatttggaagagttcagaccgtctggtcggcactgcaa 799
           || ||  | |    | |||||||| | |  |||||| | | |||| | | |||||||  ||||||
Db   16446 TTCTTTATTTAAAGAATTTGATTTAGTATCTTTTCGAACGGTTCCTCCCATACACTCCC 16505

Qy     800 gcatgcgtttggctgtgcgctggtgcacacctctggctggaaagtggtctattccgggga 859
           || |  | |||||||  |||| || || ||||| || ||||  ||||||  || |||  ||
Db   16506 GTATTCATATATTGCATGGAGATTACTAATTCCAGTGGTTGGAAGATCGCCATACTCTGGAGA 16565

Qy     860 caccatgccctgcgcgaggctctggtccggatggggaaagatgccacctccctgatacatga 919
```

```
Db  16566 TACTCGACCTTCTGAGGATTTTTGCTTAACATTGCCAAAGACTCAACTTATTTATTTATTCACGA 16625
          |||| |||||| ||||  |||||||   |||  |||||||||||  ||||
Qy    920 agccaccctggaagatggtttggaggagggaaagacaacacacagcaacgtc 979
Db  16626 AGCTACTTTGGAAGACTCTATGCCACGAAAATTGCAATAAAAAAAGCAACACAGTACTTACTC 16685
          |||||||||| |||||||
Qy    980 ccaagccatcaagcgtgg 996
Db  16686 TGAAGCTTTAGAAGTGG 16702
```

RESULT 12
ACO17383

LOCUS   ACO17383   115873 bp   DNA   HTG   09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   ACO17383
VERSION   ACO17383.1 GI:6553603
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE   fruit fly.
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
COMMENT   This sequence was identified as CDM:1021051 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
REFERENCE   1 (bases 1 to 115873)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES   Location/Qualifiers
     source   1..115873
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT   31582 a   25515 c   25855 g   33121 t
ORIGIN

Query Match   7.1%; Score 70.6; DB 64; Length 115873;
Best Local Similarity   49.2%; Pred. No. 1.1e-08;
Matches 346; Conservative   0; Mismatches 329; Indels   28; Gaps   5;

```
Qy    299 taccagaaatcatcttcctggaacaacaggtctgccatccgatgaagatcgaaatgtc 358
          |||| |||| ||  ||||||  ||||||| |||| | |||| |||    ||||| ||
Db  63659 TACCCGAAGATAATTTTCTTCGGCAACAGGCTCCGCACAGCGCCGAAACGTG 63718
Qy    359 agtgccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgaagggc 418
          ||| |||  ||| ||||| ||  ||  |||||||| | ||| |||| | |||   ||| |
Db  63719 AGCTCCATTCTGATCAGGACTGCAATTGATCCCTACGTGTTGTGTTGTGTTGCGGAGAAGGA 63778
Qy    419 acacttgggcagctgtgccgcgtcattacggagaccag--gtggacaaggtgtcctgggacc 475
          ||  ||||  |||  |||  | ||||| ||||| ||||   |||||||| |  ||||| ||
Db  63779 ACTTATATGGTCAAATTGCTACGACTATATGGCACGAGAAAAGGGCAGTTGATTCTTCCCAG 63838
Qy    476 ctggctgctgtgtttgtgtcccaacctggcacgcagaatccacacgggcttgccaagtat 535
          |||| |||| ||| | |||  |||  ||||||  |||||  ||| |||| ||||  |||
Db  63839 CTACAGGCGATTTATGTATATCCCA-TTTGCATGCCGACCACCAATTGGATTGGTCT 63897
Qy    536 cttgctgtcagagagaacgcgcctggcatctctttgggaaagccgcttcaccccttgctgt 595
          |||||   |||| |||| |  |||||  ||| |||||||||  ||| | || |||| ||||
Db  63898 GCTTCCGGAAG-----------CAGGCAATTAAAACCTAGGGAGACCCCACTTATTCT 63945
Qy    596 ggttgccccccaaccagtcaaagcctgccagacaccaacaccagtgcaggagt 655
          | ||||| || ||  || ||||  ||||| ||  |||||  ||| || |||||
Db  63946 GCTCGGCTCCTCCGCCGCCAAATTGAACCTTGGTT---GGAGTTCTACACACATTCGACAAATTAGAAAC 64002
Qy    656 cctgaccaatcacgtgtatgattcctgccaaatgcttcaggaaggggctgagagatccag 715
```

```
Db  64003 CGTAGAAGAATGGCGTATACTTCTTAGTAGTGGGCGAACGGTGAAACTCCTAGGTCCCTGAGCGG 64062
          |||| |||| | |   | ||||  |||  |||||| || || ||||  |||| |  ||| || ||
Qy    716 tcctgcagtggaagaggatgatcagttcgctgttgcgaacaatgtgattggaagagttca 775
Db  64063 TGAACAAGTTGAACGTCTAGGAATTACGTCCAATATCCACCTGCCTAGTTAGGCACNCGCC 64122
          ||| |||||||| | |  ||| ||| ||||  | |||| || || |  || ||   |
Qy    776 gacctgtctgtgtgcgggcactgcaagcatgcgttggctgtggcgctggtgtgcacacctctgg 835
Db  64123 CAACTCTTTCGGAATAAGCCTAACTCTGGCG----------GCAAAGCACAAATAGCGAACC 64173
          |||| |||| | |||  |||||| || |||           |||| |||| || |  |||
Qy    836 ctggaaagtggtctattccggggaacaccatgccctgcgaggtctgttcggatgggggaa 895
Db  64174 CGTCAAGATCACHTACACGCGCCACACHTATGGCCATHGTGTAGGACTGATCGATCTGGGCCG 64233
          || | |||| || ||||| || ||||  |||| |  |||| |||| |||| |  |||| ||
Qy    896 agatgccaccctcctgatacatggaagccaacccctggaagatggttttggaaggaagacagt 955
Db  64234 CGACTCCACCGTTCTTATTCACGAGCGGACAATGGAGGGATGACTTGGAGGAGGAGGCGCG 64293
          || | ||||  | ||||  |||| |||  |||  |||||  |  ||||| || ||| | ||
Qy    956 ggaaaaagacacacagccacacaacgtccccaagccatcagcgtgggg 998
Db  64294 GTTAAAAGACGCACAGCACCGTGTCGCAGGCCTATTCAGCGAGGGG 64336
```

RESULT 13
ACO07417/c

LOCUS   ACO07417   175682 bp   DNA   INV   02-MAR-2001
DEFINITION Drosophila melanogaster, chromosome 2R, region 47A-47B, BAC clone
            BACR48F07, complete sequence.
ACCESSION   ACO07417
VERSION   ACO07417.4 GI:13184056
KEYWORDS   HTG.
SOURCE   fruit fly.
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 175682)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.R.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE   Sequencing of Drosophila chromosome 2R, region 47A-47B
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 175682)
  AUTHORS   Celniker,S.E., Agbayani,A., Acaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory,MS 64-121, Berkeley, CA 94720, USA
COMMENT   On Mar 2, 2001 this sequence version replaced gi:5670592.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.

```
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://www.fruitfly.berkeley.edu.
            to bdgp@fruitfly.berkeley.edu.) or send email

FEATURES             Location/Qualifiers
     source          1..175682
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="47A-47B"
                     /clone="BACR48F07 (D625)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    49836 a   38215 c   38034 g   49597 t
ORIGIN

Query Match      7.18;    Score 70.6;   DB 4;   Length 175682;
Best Local Similarity   49.2%;    Pred. No. 1e-08;
Matches 346;  Conservative  0;  Mismatches 329;  Indels 28;  Gaps  5;

Qy       299 taccagaaatcatcttccttggaacaggggtctgccatcccgatgaagattcgaaatgtc 358
             |||| | || ||| ||| || |    |||||| |        |||| ||| ||| |||||
Db    108897 TACCGGAAGATAAATTTTCTTGGGCACAGGCTCCTCGCATTCCAATAAGACGGAAACGTG 108838

Qy       359 agtgcacacttgtcaacataagcccgaacgtctctgctactggactgtggtgagggc 418
             || ||    || |||  | ||  |||| ||| | || |||||  | ||| | ||  |  ||
Db    108837 AGCTCCAATTCTGATCAGGACTGCAATTGATGCCTACGTGTGTTGATTGTGGAGAAGGA 108778

Qy       419 acatttgggacgctgtgccgtcattacggaggagaccag---gtggacaggtcctgggacc 475
             || ||| |||  | | |   |   ||| |  | |  |||      || |  |||| |  ||
Db    108777 ACTTATGGTCAAAATTGTACGACTATATAIGGGACACGAGGAAAAGGCGCAGTTGAITTCTTCGCCAG 108718

Qy       476 ctggctgctgtgttgtccacctggcacggagatcaccacacggggcttgcaagtat 535
             || || ||| ||| ||| |  |||| ||| | || |  ||| || |||  ||  |
Db    108717 CTACAGGCGGATTTATGTAICCCA-TTTGCATGCCGACCACACATTGGATTGAITGGTCT 108659

Qy       536 cttgctgcaggagaacgcgcttggcatcttttgggaaagccgcttcaacccttgctggt 595
             |||  ||  |  ||  |   ||| |           ||||||  || |||  || |  |
Db    108658 GCTCCGGGAAAG---------GAGGCAATTAAAACCTAGGCAGCAGACCCACACTTATTCT 108611

Qy       596 ggttgcccccaaccagctcaaagctgctccagcaggatgccaccaacagtgcaggagt 655
             || || ||||   |  ||  ||| |   ||| |  | ||  ||  |||   | |  |
Db    108610 GCTGGGCTTCCTCGCCCCAAATTGAACCTTGGTT---GGAGTTCTACAATCGCACAAATAGAAAC 108554

Qy       656 cctgcaccatcaagtcatgatccttgcaaatgccttcaggaagggggctgagatctccag 715
             || || |  |  |  | |   ||| ||  || |  ||| | |  |  | | |
Db    108553 CGTAGAAGAATGGCGTATACTCTAGTGGGCGCAACGGTGAACTCCTAGCTACTCCCCTCGAGCGG 108494

Qy       716 tcctgcagtggaaagattgatcaagttcgtgttgcgaacatgtgatttggaagagttca 775
             |    |   | ||| |  | ||  | ||  |||   | |  |  |||  |||| |  | |
Db    108493 TGAACAAGTTGAACGTCTAGGAATTAACGTCATTCACCTGCCTAGTTAGGGCACTGCCC 108434

Qy       776 gacctgtctggtgcggcactgaagacatgcgttggctgcacacctctgg 835
             | ||| | ||  |  ||||   ||  |    |||  |||  | || |||
Db    108433 CAACTCTTTCGGAATAAGCCTAACTCTGGCG----------GCAAAGCACAATAGCGAACC 108383

Qy       836 ctggaaagtggtctattccggggaacacatgccctgcgaggtctctgtccggatgggaa 895
             |||  |  || |  ||| ||||| | |  |  | ||||| |||  |||| ||  | | |
Db    108382 CGTCAAGATCACTTACACGGCGGACACTATGCCATGTCAGGATCTGACGAICTGGGCCG 108323

Qy       776 agatgccaccctcctgatacatggaagagtgttcttggaaggaagaagcagt 955
             |   |  ||   ||| |   |  | ||   ||  | ||  | |  |  | ||
Db    108382 CGTCAAGATCACTTACACGGCGGACACTATGCCATGTCAGGATCTGACGAICTGGGCCG 108323

Qy       896 agatgccaccctcctgatacatggaagagtgttcttggaaggaagaagcagt 955
             |   |  ||   ||| |   |  | ||   ||  | ||  | |  |  | ||
Db    108322 CGACTCCACCGTTCTTAITTCACCAGGCCCACACAAIGGAGGATGACTTGGAGGAGGCCCG 108263

Qy       956 ggaaaaggacacacaagcacacaaccatcagcgtgggg 998
             |   ||| |  |  | || |  ||||  ||  | ||||
Db    108262 GTTAAAGACGACAGCACGCACCGGTGTCGCAGGCTATTCACCAGGGG 108220

RESULT   14
```

LOCUS   AC007352   190574 bp   DNA   HTG   24-FEB-2000
DEFINITION Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPCI-98
            19.J.16 map 47A-47B strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 36 unordered pieces.
ACCESSION   AC007352
VERSION   AC007352.8 GI:7025687
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   fruit fly.
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 190574)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE   Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 190574)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT   On Feb 24, 2000 this sequence version replaced gi:7018750.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://www.fruitfly.berkeley.edu./) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 36 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1   678: contig of 678 bp in length
            *   679   758: gap of unknown length
            *   759   1358: contig of 600 bp in length
            *   1359   1438: gap of unknown length
            *   1439   2373: contig of 935 bp in length
            *   2374   2453: gap of unknown length
            *   2454   2962: contig of 509 bp in length
            *   2963   3042: gap of unknown length
            *   3043   3889: contig of 847 bp in length
            *   3890   3969: gap of unknown length
            *   3970   5437: contig of 1468 bp in length
            *   5438   5517: gap of unknown length
            *   5518   5572: contig of 1055 bp in length
            *   5573   6652: gap of unknown length
            *   6653   8387: contig of 1735 bp in length
            *   8388   8467: gap of unknown length
            *   8468   9060: contig of 593 bp in length
            *   9061   9140: gap of unknown length
            *   9141   11167: contig of 2027 bp in length
            *   11168   11247: gap of unknown length
            *   11248   14518: contig of 3271 bp in length
            *   14519   14598: gap of unknown length
```

```
Qy    359 agtgccaacacttgtcaacataagcccgacaacgtctctgctactggactggtggtgagggc    418
          |||||  ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db 132812 TACCCGAAGATAATAATTTTCTTGGGCACAGGCTCCTCCATTCCCAATAAGACGCGAAACGTG 132871
Qy    299 taccagaaatcatcctccttggaacaaggtgctgccatcccgatcgaagattcgaaatgtc    358

Matches 346;  Conservative   0;  Mismatches 329;  Indels   28;  Gaps    5;
Best Local Similarity    49.2%;    Pred. No. 1e-08;
Query Match        7.1%;    Score 70.6;   DB 60;   Length 190574;

BASE COUNT    53370 a   39726 c   40358 g   54299 t   2821 others
ORIGIN
```

FEATURES   Location/Qualifiers
     source   1..190574
                     /organism="Drosophila melanogaster"
                     /strain=y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="47A-47B"
                     /clone="BACR19J16"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"

```
*      1459    18837:  contig of 4239 bp in length
*      18838   18917:  gap of unknown length
*      18918   24324:  contig of 5407 bp in length
*      24325   24404:  gap of unknown length
*      24405   29984:  contig of 5580 bp in length
*      29985   30064:  gap of unknown length
*      30065   35794:  contig of 5730 bp in length
*      35795   35874:  gap of unknown length
*      35875   44808:  contig of 8934 bp in length
*      44809   44888:  gap of unknown length
*      44889   58222:  contig of 13334 bp in length
*      58223   58302:  gap of unknown length
*      58303   78322:  contig of 20020 bp in length
*      78323   78402:  gap of unknown length
*      78403   122249:  contig of 43847 bp in length
*      122250  122329:  gap of unknown length
*      122330  179102:  contig of 56773 bp in length
*      179103  179182:  gap of unknown length
*      179183  179407:  contig of 225 bp in length
*      179408  179487:  gap of unknown length
*      179488  180592:  contig of 1105 bp in length
*      180593  180672:  gap of unknown length
*      180673  181383:  contig of 711 bp in length
*      181384  181463:  gap of unknown length
*      181464  182092:  contig of 629 bp in length
*      182093  182172:  gap of unknown length
*      182173  183480:  contig of 1308 bp in length
*      183481  183560:  gap of unknown length
*      183561  183771:  contig of 211 bp in length
*      183772  183851:  gap of unknown length
*      183852  184058:  contig of 207 bp in length
*      184059  184138:  gap of unknown length
*      184139  185035:  contig of 897 bp in length
*      185036  185115:  gap of unknown length
*      185116  185846:  contig of 731 bp in length
*      185847  185926:  gap of unknown length
*      185927  186560:  contig of 634 bp in length
*      186561  186640:  gap of unknown length
*      186641  187253:  contig of 613 bp in length
*      187254  187333:  gap of unknown length
*      187334  187611:  contig of 278 bp in length
*      187612  187691:  gap of unknown length
*      187692  188409:  contig of 718 bp in length
*      188410  188489:  gap of unknown length
*      188490  189196:  contig of 707 bp in length
*      189197  189276:  gap of unknown length
*      189277  189969:  contig of 693 bp in length
*      189970  190049:  gap of unknown length
*      190050  190574:  contig of 525 bp in length.
```

```
REFERENCE   1  (bases 1 to 261846)
AUTHORS     Adams,M.D.,  Celniker,S.E.,  Holt,R.A.,  Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G.,  Scherer,S.E.,  Li,P.W.,  Hoskins,R.A.,  Galle,R.F.,
            George,R.A.,  Lewis,S.E.,  Richards,S.,  Ashburner,M.,  Henderson,S.N.,
            Sutton,G.G.,  Wortman,J.R.,  Yandell,M.D.,  Zhang,Q.,  Chen,L.X.,
            Brandon,R.C.,  Rogers,Y.H.,  Blazej,R.G.,  Champe,M.,  Pfeiffer,B.D.,
            Wan,K.H.,  Doyle,C.,  Baxter,E.G.,  Helt,G.,  Nelson,C.R.,  Gabor
            Miklos,G.L.,  Abril,J.F.,  Agbayani,A.,  An,H.J.,
            Andrews-Pfannkoch,C.,  Baldwin,D.,  Ballew,R.M.,  Basu,A.,
            Baxendale,J.,  Bayraktaroglu,L.,  Beasley,E.M.,  Beeson,K.Y.,
            Benos,P.V.,  Berman,B.P.,  Bhandari,D.,  Bolshakov,S.,  Borkova,D.,
            Botchan,M.R.,  Bouck,J.,  Brokstein,P.,  Brottier,P.,  Burtis,K.C.,
            Busam,D.A.,  Butler,H.,  Cadieu,E.,  Center,A.,  Chandra,I.,
            Cherry,J.M.,  Cawley,S.,  Dahlke,C.,  Davenport,L.B.,  Davies,P.,  de
            Pablos,B.,  Delcher,A.,  Deng,Z.,  Mays,A.D.,  Dew,I.,  Dietz,S.M.,
            Dodson,K.,  Doup,L.E.,  Downes,M.,  Dugan-Rocha,S.,  Dunkov,B.C.,
```

```
REFERENCE
AUTHORS

ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
SOURCE      fruit fly.
KEYWORDS    HTG.
VERSION     AE003830.2  GI:10727666
ACCESSION   AE003830  AE002787
            of 52, complete sequence.
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386047 section 13
LOCUS       AE003830    261846 bp    DNA        INV       04-OCT-2000
```

```
RESULT  15
AE003830/c
```

```
Db 133447 GTTAAAAGACGCACAGCACCGTGTCGCAGGCTAATTCAGCAGGGG 133489
          |||||||||  ||||||  ||  |||||||  ||  |||||||||
Qy    956 ggaaaagaacacacagcacaagtcccaagccaatcaagcgtgggg 998
Db 133387 CGACTCCACCGGTTCTTATTCACGAGCGACAACAATGGGAGGATGGACTTGGAGGAGGCGCCG 133446
          | |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Qy    896 agatgccacctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagt 955
Db 133327 CGTCAAGATACACTTACACGCGGCGACACTATGCCATTGTCAGGATCTGATCGATCTGGCCG 133386
          || |||||||  ||| |||| |||| |||||||| |||||  |||||||||  |||||
Qy    836 ctggaaagtggtctattatccgggaacaccatgccctgccgaagcgctctgtccggatgggaa 895
Db 133276 CAACTCTTTCGGAATAAGCCTAACTCTGCCG----------GCAAAGACAAATAGCGGAACC 133326
          |||||  ||||||||  | |||| |||||||          | |||  |||| |||| |||
Qy    776 gacctgtctgtgcggcactgcaagacatgcttgtccggtcgtcgtgtgtgtgcacacctgg 835
Db 133216 TGAACAAGTTGAACGTCTAGGAATTACGTCCAATATCCACTCGGTTAGCCACTCGCC 133275
          ||||| |||| |||||  |||||||  ||  ||||  |||  ||||||||  ||
Qy    716 tcctgcagtggaaagattgatcagttcgctgttcgacacatgtgatttggaaagttca 775
Db 133156 CGTAGAAGATGCGTATACTTCAGTGTGGGCGCAACGGTGAACTCCTAGCTTAGTCCCTGACCGG 133215
          |||| |||||||  ||| || || |||||  ||| ||||||||||||  ||||  || ||
Qy    656 cctgcacacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccag 715
Db 133099 GCTGGCTCCTCCGCCAAATTGAACCTTGGGTT---GGAGTTCTACAATCGACAAATAGAAAC 133155
          |||||| || |||| ||||||||||  ||||   ||||||| ||||||  ||||  |||
Qy    596 ggttgccccacaagtcaaagcttccagcagtacccacacaacagtgcagaggt 655
Db 133051 GCTCCCGGAAAG--------GAGGCAATTAAAACCTAGGGCAGAGCACAGACCCCACTTATTCT 133098
          |||||||||||        |||| ||||||||  | |||| ||| |||  ||  || |
Qy    536 cttgctgcaggagaacgcgccttgggaaagccgcttcacctttgtggt 595
Db 132992 CTACAGGCGATTTATGTATCCCA-TTTGCATGCCGACCACCACATTGGATTGATTGGTGT 133050
          ||| || || ||| || ||| |  ||| ||| |||| |||  |||  || ||| |
Qy    476 ctggctgctgtttgtgtccacctgcacgcaagatcaccacacggcttgccaagtat 535
Db 132932 ACTTATAGGGTCAAATTGTACGACTATATGTCACGAGAAAAGGGCAGTTGAATTCTTGCCAG 132991
          ||| ||||||||||||| ||  |||| ||||| |||  |||||||||  |||| ||||
Qy    419 acattgggcagctgtgcctgtcattacggacgaccag--gtggacagggtcctgggcacc 475
Db 132872 AGCTCCANTCTGATCACGAGACTGCAAATGATGCCTACGTGTTGTTGGTGGATTGTGGAGAAGGA 132931
```

```
           Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
           Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
           Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
           Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
           Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
           Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
           Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
           Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
           Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
           McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
           Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
           Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
           Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
           Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
           Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
           Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
           Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
           Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
           Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
           Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
           Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
           Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
           Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE    The genome sequence of Drosophila melanogaster
  JOURNAL  Science 287 (5461), 2185-2195 (2000)
  MEDLINE  20196006
REFERENCE   2  (bases 1 to 261846)
  AUTHORS  Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE    Direct Submission
  JOURNAL  Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
           Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7303755.
FEATURES             Location/Qualifiers
     source          1..261846
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            join(960..1069,1135..1237,1290..1357,1416..1731,
                     2658..2759,3801..4142,4200..4354,4727..4835,
                     6499..6941,6999..7841,7921..8315)
                     /gene="Pfk"
                     /product="CT13302"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     gene            <960..>8315
                     /gene="Pfk"
                     /note="pfk"
                     /map="46E3-46E6"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     CDS             join(1026..1069,1135..1237,1290..1357,1416..1731,
                     2658..2759,3801..4142,4200..4354,4727..4835,
                     6499..6941,6999..7841,7921..8059)
                     /gene="Pfk"
                     /note="Pfk gene product [alt 1]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
                     /protein_id="AAF58840.1"
                     /db_xref="GI:7303792"
                     /translation="MHSIKFRVETKLKPIFLEINGRIPICRHHGPTPFRLEISNKTP
                     PIRQKLFTPNIGLTQCTRSHHLCCPRDISGNTLLSVKFNCKHCHLKLRDSGDQKNDSP
                     GEKNIOKRKSAQRCGKPINNLHNGFLNAVNYSEKNAVKKHKSAPKRKCGKSVDELRKC
                     LRTMQDVIDFVHPVKPFKDKGLAVFTSGGDSQGMNAVARACVRMAILGCKVFIREG
                     YQGMVDGGDCIQEANWASVSSIIHRGGTIGSARCQDFREROGRLKAANNLIORGITN
                     LVVIGGDSSLTGANLFRQEWSSLLDELVKNKTITTEQOEKFNVLHIVGLVGSIDNFC
                     GIDMTIGTDTALHRIIEAIDAISSTAYSHQRTIMEVWGRHCGYLALVGGLACEADFI
                     FIPEMPKPKVDWPDRLCSQLAQERSAGQRLNIVIVAEGAMDREGHPITRAEDVKVIDER
                     LKHDARIIVLGHVORGGNPSAFDRILACRMGAEATLALMEATKDSVPVVISLDGNQAV
                     RVPLMECVERTQAVAKAMAEKRWADAVKLRGSFERNLETYKMLTRLKPPKENFDADG
                     KGIEGYRLAVMHIGAPACGMNAAVRSFVRNAIYRGDVVYGINGGVEGLIAGNVRELGW
                     SDVSGWVCQGGAYLGTKTKRFLPEGKFKETAARLKEFKIQGLLIGGFESYHAAGQIADQ
                     RDNYPQFCIPIVVIPSTISNNVPGTEFSLGCDIGLNEITEICDRIRQSAGTKRRVFV"
     mRNA            join(2152..2759,3801..4142,4200..4354,4727..4835,
                     6101..6289,6499..6941,6999..7841,7921..8315)
                     /gene="Pfk"
                     /product="CT13275"
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
     CDS             join(2613..2759,3801..4142,4200..4354,4727..4835,
                     6101..6289,6499..6941,6999..7841,7921..8059)
                     /gene="Pfk"
                     /note="Pfk gene product [alt 2]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004001"
                     /db_xref="FLYBASE:FBgn0003071"
                     /protein_id="AAF58841.1"
                     /db_xref="GI:7303793"
                     /translation="MNSEINQRFLARGSOKDKGLAVFTSGGDSQGMNAVARACVRMAI
                     YLGCKYFIREGYQGMVDGGDCIQEANWASVSSIIHRGGTTIGSARCQDFREROGRLK
                     AANNLIORGITNLVVIGGDGSLTGANLFRQEWSSLLDELVKNKTITTEQOEKFNVLHI
                     VGLVGSIDNFCGTDMTIGTDTALHRIIEAIDAISSTAYSHQRTIMEVWGRHCGYLA
                     LVGGLACEADFIFIPEMPKPKVDWPDRLCSQLAQERSAGQRLNIVIVAEGAMDREGHPI
                     TAEDVKKVIDERLKHDARIIVLGHVORGGNPSAFDRILACRMGAEATLALMEATKDSV
                     PVVISLDGNQAVRVPLMECVERTQAVAKAMAEKRWADAVKLRGSFERNLETYKMLTR
                     LKPPKENFDADGKGIEGYRLAVMHIGAPACGMNAAVRSFVRNAIYRGDVVYGINDGVE
                     GLIAGNVRELGWSDVSGWVCQGGAYLGTKTKRFLPEGKFKETAARLKEFKIQGLLIGGF
                     ESYHAAGQIADQRDNYPQFCIPIVVIPSTISNNVPGTEFSLGCDIGLNEITEICDRIR
                     QSAQGTKRRVFVIETMGGYCGYTLATLLAGLAGGADAAYIYEEKFSIKDLQQDVYHMASK
                     MAEGVSRGLLLRNEKASENYSTDFIYRLYSEEGKGLFTCRMNILGHMQQGGSPTFPDR
                     NMGTKMAAKCVDWLAAQIKANIDANGVVNCKSPDTATLLGIVSRQYRFSPLVDLIAET
                     NFDQRIPKKQWWLRLRPLLRILAKHDSAYEEGGMYITVEEBCDTDAVA"
     mRNA            join(8915..9052,9216..9726,10663..10911,10976..11174,
                     11235..11542,11596..>12210)
                     /gene="CG11866"
                     /product="CT37014"
                     /db_xref="FLYBASE:FBan0011866"
                     /db_xref="FLYBASE:FBgn0033486"
     gene            <8915..>12210
                     /gene="CG11866"
                     /db_xref="FLYBASE:FBan0011866"
                     /db_xref="FLYBASE:FBgn0033486"
     CDS             join(9286..9726,10663..10911,10976..11174,11235..11542,
                     11596..12210)
                     /gene="CG11866"
                     /note="CG11866 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0011866"
                     /db_xref="FLYBASE:FBgn0033486"
                     /protein_id="AAF58839.1"
                     /db_xref="GI:7303791"
                     /translation="MYSTRQMCSGGGTSASEGVSPFGAGSGGSASGTGMASGSAVSSG
                     SGTGVOEPPESTANRSSVLRTTSYQGNQAHSSSYRVVSAGSLAESASAVIRLHSYTMQ
                     MQRORPPTATTTTTSAPSIYDLPNELIEKILSYVDYKKVSNLRLVSHRMNDICMAML
                     NTAFTKQIKTTLSRFQAIKASMPRRSHRNHPLACECDIIETCYMRLSLLQMSMGKH
                     IERGHCCFPGAILDEVQAILNYISITPRLQRPYRVTDELFDLSTMAMEYFKDRIEAT
                     LPGLAYFNKDFYTLPTTFKRPTLAISSDLEDSASNSPPQNHMVLRRGIRKIIKQGMKMY
                     NNQLSVLRTELRTCRKRAAEQSKOLAEOSNIISEQOKOTLEYANRLDENDKKNEEMSR
                     KFSTLLQELNKCKTELQFWRSKSPAIPAVCSSCNQKVAPVPWPPEDEQALYNQGVDPEH
                     FIIINEDVDAESDSVSGELKEFASPDESTFAKLLAVSTAANRMENKVKNLRNDICMAML
                     TSKAAEVAQSOSOPMATGNAGAAAKAAGGYSPKLFYQNHQRSGVISPVSQNLGVAATT
                     DNVAPGSEALEEPEAKKARRVQKANRYVNTPGKRSK"
     complement(join(<12115..12937,12999..13642,13694..13708))
                     /gene="CG17753"
                     /product="CT11457"
                     /db_xref="FLYBASE:FBan0017753"
                     /db_xref="FLYBASE:FBgn0033487"
     gene            complement(<12115..>13708)
                     /gene="CG17753"
                     /map="46E4-46E6"
                     /db_xref="FLYBASE:FBan0017753"
```

```
CDS        /db_xref="FLYBASE:FBgn0033487"
           complement(join(12802. .12937,12999. .13642,13694. .13708))
           /gene="CG17753"
           /note="CG17753 gene product"
           /codon_start=1
           /db_xref="FLYBASE:FBan0017753"
           /db_xref="FLYBASE:FBgn0033487"
           /protein_id="AAF58838.1"
           /db_xref="GI:7303790"
           /translation="MSSIKIEFAVQMRRGDESYAGALRSALDGVGQVEIDTQGRVII
           QTQRPWSEIQDKIEATGVRAVLSGPGGQSAVALINTTGSVVDKTPIQGVVRFTITAD
           KKPGVVVDGVGVDGLSPGLHGLHIHESGDTSAGCSSVGEHYNPRQSPHGSPAAGAEERH
           AGDLGNIRADENGRATFRFRVDVPLEVWDIIGRAVVLTANADDLGRGGNDQSLIDGNSG
           ERIACGIIARSAGIIENFKRICACDGVTLWDERNKPLAGKERSQKL"
mRNA       join(<18636. .18761,19765. .19876,19968. .20213,22806. .22878,
           22958. .23344,23489. .23633,23934. .24407,25026. .>25202)
           /gene="CG11867"
           /product="CT37016"
           <18636. .>25202
           /db_xref="FLYBASE:FBan0011867"
           /db_xref="FLYBASE:FBgn0033488"
gene       /gene="CG11867"
           /map="46E4-46E8"
           /db_xref="FLYBASE:FBan0011867"
           /db_xref="FLYBASE:FBgn0033488"
CDS        join(18636. .18761,19765. .19876,19968. .20213,22806. .22878,
           22958. .23344,23489. .23633,23934. .24407,25026. .25202)

Query Match        7.1%;  Score 70.6;  DB 5;  Length 261846;
Best Local Similarity  49.2%;  Pred. No. 1e-08;
Matches 346;  Conservative  0;  Mismatches 329;  Indels  28;  Gaps  5;

Qy    299 tacccagaaatcatcctccttggaacaggtgtctgccatccccgatgaagatcgaaatgtc 358
          ||||||  || || |||| |  ||| ||| |||||||||||| | |||||||| | |||||
Db 199291 TACCCGAAGATAATAATTTTCTTGGGCACAGGTGTCCCTCCGATTCCCAATAAGACGGAAACGTG 199232

Qy    359 agtgccacacttgtcaacataagcccgacacgtctctgctactggactttggtgaggc 418
          || |||  || || || |  ||  ||| || | | |  ||| | |||||||| | |||
Db 199231 AGCTCCATTCTGATCAGGACTGCAAATTGATGCCTACGTGTGTTGTTGGATTTGTGGAAGGA 199172

Qy    419 acatttgggcagctgtgccgtcattacggagaccag---gtggacagggtcctgggcacc 475
          | | |||  ||||||  | || |  || ||| |||     |||| ||| |||| |||||
Db 199171 ACTTATGGTCAAAATTGTACGACTATATGGACACGAGAAAGGCAGTTGATTCTTCGCCAG 199112

Qy    476 ctggctgctgtgtttgtgtccacctggcacgcagatcaccacacgcggcttgcaagtat 535
          || | |||  || ||| ||||  | ||||  || ||| | || | ||| |||  ||||||
Db 199111 CTACAGGCGGATTTAGTGTATCCCA-TTTGCAGCCCGACCACACATTGGATTGATTGGTCT 199053

Qy    536 cttgctgcaggagaaacgccctggcaatctttgggaaagccgcttcaccctttgctggt 595
          ||| ||| || || | | || |   || |||||||||||| | || || || |  ||||
Db 199052 GCTCCGGGAAAG-------GAGGCAATTAAAACCTAGGGCAGACCCACTTATTCT 199005

Qy    596 ggttgcccccaaccagctcaaagcctggctccagtaccacacacagtgccaggaggt 655
          || |||  ||||||| ||||| ||  |||||||| || | ||||| || |||||||
Db 199004 GCTGGCTCCTCCGCCAAAATTGAAACCTTGGTT---GGAGTTCTACAATCGACAAAATAGAAAC 198948

Qy    656 cctgcaccacatcagtatgattcctgccaaatgccttcaggaagggggctgagatccag 715
          || ||||| | | | | |  || ||||  ||  ||| ||| || | |||| |||| |||
Db 198947 CGTAGGAAGATGCGTATACTCTAGTGGGCAACGGTGGAACTCCTAGCTAGTCCCCTGAGCCG 198888

Qy    716 tcctgcagtggaaagattgatcagttcgctgttgccgaaacatgtgattttgaaagagtttca 775
          |||  ||| || |  || | |||  |  | ||| ||||| ||| | |||  || ||||
Db 198887 TGAACAAGTTGAACGTCTAGGAATTACGTCCATATCCACCTGCCTAGGTTAGGCCACTGCCC 198828

Qy    776 gacctgtctggtgcggcactgcaagacatgcgttggctgtgcgctggttgcacacctctg 835
          | ||  |||| | |  |||   ||| ||| ||||||||| | |  | ||| | |||
Db 198827 CAACTCTTTCGGAATAAGCCTAACTCTGGGG--------GCAAAGCACAATAGGCAAACC 198777

Qy    836 ctggaaagtggtctattccgggggaccaccacctgcctcgccgggatgggaa 895
          ||| ||||| || || ||||||||  || || | ||| || |  |||| ||
Db 198776 CGTCAAGATCACTTACAGGCGCGACACTATGCCATGTCAGGATCGATCTGGGCCG 198717

Qy    896 agatgccaccctcctgatacatgaagcacccctggaaagatggtttggaagaggaagcagt 955
          . | || || ||| ||  ||  || ||| ||||| |||| |||||| |||||| || |||
Db 198716 CGACTCCACCGTTCTTATTCACGAGCGCGACAATGGAGGATGACTTGGAGGAGGAGGAGGCGCG 198657

Qy    956 ggaaaagacacaagcaacagtcccaagccatcaagcgtgggg 998
          |||||||| || |||| || |  |||||||||  |  |||||
Db 198656 GTTAAAGACGCGACAGCACCGTGTCGCAGGCTATTCAGCAGGGG 198614

Search completed: October 28, 2001, 21:51:39
Job time: 6372 sec
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: October 28, 2001, 21:04:27 ; Search time 100.36 seconds
                   (without alignments)
                   6262.749 Million cell updates/sec

Title: SEQLINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacaagattcaaa...........aagccatcagcgtgggatg 1001

Searched: 730101 seqs, 313950809 residues
Scoring table: Gapop 10.0 , Gapext 1.0
                   IDENTIfY_NUC

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_0601:*

c   1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
c   2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
c   3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
c   4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
c   5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
c   6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
c   7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
c   8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
c   9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
c 10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
c 11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
c 12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
c 13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
c 14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
c 15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
c 16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
c 17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
c 18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
c 19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
c 20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
c 21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
c 22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	989	98.8	2546	21	AAC76445	Human ORFX ORF2000
2	989	98.8	2958	21	AAA58453	Human prostate can
3	984.2	98.3	2478	21	AAA58810	Human sulphatase G
4	136.6	13.6	26664	21	AAA60207	Human prostate can
5	70	7.0	1617	21	AAC34173	Arabidopsis thalia
6	70	7.0	2853	21	AAC49985	Arabidopsis thalia
7	41.6	4.2	464	21	AAC38171	Arabidopsis thalia
8	37.6	3.8	521	21	AAC04378	Human secreted pro
9	37	3.7	2517	22	AAC89621	S. cerevisiae YKR0
10	36.4	3.6	10732	21	AAA10594	Gene encoding a su
11	36	3.6	1409	21	AAC42402	Arabidopsis thalia

ALIGNMENTS

RESULT   1
AAC76445
ID   AAC76445 standard; cDNA; 2546 BP.
XX
AC   AAC76445;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999; 99US-0127607.

c 12   35.8   3.6   19053   21   AAA59146   Nucleotide sequenc
c 13   35.4   3.5   366   20   AAV89531   EST clone CP314.
c 14   35.2   3.5   1716   22   AAF26315   Pseudomonas sp lip
c 15   35   3.5   3474   14   AAQ39018   Sequence of pure m
c 16   35   3.5   6901   18   AAT78853   Human lecithin-cho
c 17   35   3.5   8854   20   AAA20086   DNA encoding chime
c 18   34.8   3.5   4173   21   AAA62899   Murine JNK3 bindin
c 19   34.8   3.5   4200   21   AAA62900   Murine JNK3 bindin
c 20   34.8   3.5   4266   21   AAA62901   Murine JNK3 bindin
c 21   34.8   3.5   4269   21   AAA62902   Murine JNK3 bindin
c 22   34.2   3.4   2024   11   AAQ03143   Nucleotide sequenc
c 23   34   3.4   300   20   AAA14375   Human gene express
c 24   34   3.4   2022   21   AAA49455   Human Voltage-gate
c 25   33.8   3.4   928   21   AAC77016   Human ORFX ORF2571
c 26   33.4   3.3   1032   21   AAF07546   Fusarium venenatum
c 27   33.4   3.3   1128   20   AAA06453   cDNA of Baboon Gro
c 28   33.4   3.3   3474   18   AAT74012   Maize optimised-B.
c 29   33.4   3.3   3474   19   AAV16191   Maize optimised DN
c 30   33.4   3.3   3508   14   AAQ39020   Sequence of full l
c 31   33.4   3.3   3508   18   AAT74013   Full length maize
c 32   33.4   3.3   3508   19   AAV16192   Maize optimised DN
c 33   33.4   3.3   3840   10   AAN92365   Sequence of maize
c 34   33.2   3.3   1791   19   AAV59723   Human secreted pro
c 35   33.2   3.3   1800   22   AAF61102   P. putida KT2440-a
c 36   33   3.3   569   21   AAA56065   Eucalyptus grandis
c 37   33   3.3   3483   21   AAA91976   Mouse trophinin DN
c 38   33   3.3   43804   18   AAT86375   Chicken embryo let
c 39   33   3.3   43804   20   AAX26690   Complete genome se
c 40   32.8   3.3   696   21   AAA50885   A. sulcata humanis
c 41   32.8   3.3   1401   20   AAA23803   Human Kv6.2 cDNA.
c 42   32.8   3.3   1872   11   AAQ06828   Human secreted pro
c 43   32.8   3.3   2385   18   AAF72725   Extracellular port
c 44   32.8   3.3   2871   21   AAZ50586   DC8scFv-erbB2EC fu
c 45   32.8   3.3   3600   21   AAA89736   Human HER-2/neu co

```
PR   02-APR-1999;
PR   05-APR-1999;   99US-0127636.
PR   30-MAR-2000;   99US-0127728.
XX                  2000US-0540763.
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-6023362/57.
DR   P-PSDB; AAB42236.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3179-3180; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;
```

```
Query Match             98.8%;  Score 989;  DB 21;  Length 2546;
Best Local Similarity   99.9%;
Matches 1000;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;
              Pred No. 1.1e-266;
```

```
Qy       1 tcgcagccacaagagattcaaacccagctcaactcatccaccggacactcttccctct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 tcgcagccacaagagattcaaacccagctcaactcatccaccggacactcttccctct 797

Qy      61 caccagtttccgctgtaagaagaaggaggccccaccctcagtgttgccatggttcagggtga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 caccagtttccgctgtaagaagaaggaggccccaccctcagtgttgccatggttcagggtga 857

Qy     121 atgcctcctcaagtaccagctccgtcccagaggagggggatgcattatac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 atgcctcctcaagtaccagctccgtcccagaggagggggatgcattatac 917

Qy     181 ttgcaatcctgaggaattcatagttgagcgctgcagttccaacttccagcagacgct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 ttgcaatcctgaggaattcatagttgagcgctgcagttccaacttccagcagacggt 977

Qy     241 gcaggagtacaggaggagtcgcagacgagggccagagcagagaaagaagtcagta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 gcaggagtacaggaggagtcgcagacgagggccagagcagagaaagaagtcagta 1037

Qy     301 cccagaaatcacctcttccttggaacaggtgtctgccatcccgatgaaagtgtcag 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 cccagaaatcacctcttccttggaacaggtgtctgccatcccgatgaaagtgtcag 1097

Qy     361 tgccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgaggcac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 tgccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgaggcac 1157
```

```
Qy     421 atttgggcagctgtccgtcattacggagaaggtggacaaggtggacaaggtatcttgc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 atttgggcagctgtccgtcattacggagaaggtggacaaggtggacaaggtatcttgc 1217

Qy     481 tgctgtgtttgtccacctgcacgcaggcagatcaccacaggctcaagtatcttgc 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 tgctgtgtttgtccacct-gcacgcaggatcaccacacaggcttgccaagtatcttgc 1276

Qy     541 tgcagagagaacgcgccttggccatccttcggggaaagcgatacttcaccctttgct 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1277 tgcagagagaacgcgccttggccatcttcggggaaagcgcttcaccctttgctgtggttg 1336

Qy     601 cccccaaacagctcaaagcctggctccagcagtaccacaacaggctgccagaggtcctgc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1337 cccccaaacagctcaaagcctggctccagcagtaccacaacaggctgccagaggtcctgc 1396

Qy     661 accacatcagtatgatgattcctgccaaatgcctcagagaagggggctgagatctcagtcctg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1397 accacatcagtatgatgattcctgccaaatgcctcagagaagggggctgagatctcagtcctg 1456

Qy     721 cagtggaaaagattcagttcgctgtcgttggcaaacatgtccaaacatgtgattgatttggaagacct 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1457 cagtggaaaagattgatcaagttcgctgttggcaaacatgtgattgattttggaagacct 1516

Qy     781 gtcctgtcgtcggcactgcaagcatcgtcgttttggctgtcgcgctggtgtcacacctctggctgga 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1517 gtcctgtcgtcgggcactgcaagcatcgtcgttttggctgtcgcgctggtgtcacacctctgga 1576

Qy     841 aagtggtctattccgggggacaccatgcccctgccgaaggctctggtccggatggggaaagatg 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1577 aagtggtctattccgggggacaccatgcccctgccgaaggctctggtccggatgggaaagatg 1636

Qy     901 ccaccctcctgatacatgaagccaccctggaagatggttggaggaaggaaagacagcaggaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1637 ccaccctcctgatacatgaagccaccctggaagatggttggaggaaggaaagacagcaggaaa 1696

Qy     961 agacaacacagcacaacgtcccaagccatcaagcgtcggggatg 1001
           ||||||||||||||||||||||||||||||||||||||||||||
Db    1697 agacaacacagcacaacgtcccaagccatcaagcgtcggggatg 1737
```

```
RESULT 2
AAA58453
ID   AAA58453 standard; cDNA; 2958 BP.
XX
AC   AAA58453;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..2531
FT                   /*tag= a
FT                   /product= "HPC2"
XX
PN   WO20002786A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
```

```
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Claim 3; Page 98-100; 157pp; English.
XX
CC   The present sequence is the coding sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;

Query Match     98.8%;  Score 989; DB 21; Length 2958;
Best Local Similarity    99.9%;  Pred. No. 1.2e-266;
Matches 1000;  Conservative   0;  Mismatches   0;  Indels    1;  Gaps    1;

Qy          1 tcgcagccacacaagattcaaacaaccagctcaacctcatccaccggacaatcttccccgct 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1190 tcgcagccacacaagattcaaacaaccagctcaacctcatccaccggacaatcttccccgct 1249

Qy         61 caccagttccgctgtaagaagaggagggcccccaccctcagtgtgccatggttcaggtga 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1250 caccagttccgctgtaagaagaggagggccccacccctcagtgtgccatggttcaggtga 1309

Qy        121 atgcctcctcaagtaccagctccgtcccaggaggaggatgccagcattattac 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1310 atgcctcctcaagtaccagctccgtcccaggaggaggatgccagcattattac 1369

Qy        181 ttgcaatcctgaggaattcatagttgagggcctgcagctccacttccagcagagcgt 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1370 ttgcaatcctgaggaattcatagttgagggcctgcagctcccaacttccagcagagcgt 1429

Qy        241 gcaggagtaccaggaagtgcgcaggacggacgcccagcccagcagagagaagaagtcagta 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1430 gcaggagtacagggaagtgcgcaggacggacgcccagcccagcagaaaagaagtcagta 1489

Qy        301 cccagaaatcaatctctccttggaacagggtctgccatcccggatgaagattcgaaatgtcag 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1490 cccagaaatcatctctccttggaacagggtctgccatcccggatgaagattcgaaatgtcag 1549

Qy        361 tgccacacttgtcaacataagccccaacagtctctcgctactgactgttgtgtgagggcac 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1550 tgccacacttgtcaacataagccccaacagtctctcgctactgactgttgtgtgagggcac 1609

Qy        421 atttgggcagtgtccgtcattacggagaccaggtggacaggtcctgggccaccctggc 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1610 atttgggcagctgccgtcattacggagaccaggtggacaggtcctgggccaccctggc 1669

Qy        481 tgctgtgtttgtccacaccctggagcagatcacacagggcgcttgccaagtgtatcttgc 540
              ||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||||
Db       1670 tgctgtgtttgtccacacct-gcacgcagatcacacagggcttgccaagtatcttgc 1728

Qy        541 tgcagagagaacgcgcctggcatctttgggaaagccgcttcacccttgctggttggttg 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 tgcagagagaacgcgcctggcatctttgggaaagccgcttcacccttgctggttg 1788

Qy        601 ccccaaccagctcaagctgctggcttccacagctgccaagtagcaggaggtcctgc 660
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1789 ccccaaccagctcaagctcggctccacagtaccacacaccaacaccaggaggtcctgc 1848

Qy        661 accacatcagtatgattcctgcaaatgcctcaggaagggctgaaatcccagtcctg 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 accacatcagtatgattcctgcaaatgcctcaggaagggctgaaatcccagtcctg 1908
```

```
Qy        721 cagtggaaagatggatcagttcgctgtcgtggaacatgcagtttcggttggaaagagttcaagacct 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 cagtggaaagatggatcagttcgctgtcgtggaacatgcagtttcggttggaaagagttcaagacct 1968

Qy        781 gtctggttcgcggcaagcaagcatggcagtttcggtctgtcgtgctgtgtggcacactctggtgga 840
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1969 gtctggttcgcggcaactgcaagcatggcagtttcggtctgtcgtgctgtgtggcacactctggctga 2028

Qy        841 aagtggtctattccgggacacatgcccctgccaggtctcggttcccggtggggaaagtg 900
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       2029 aagtggtctattccgggacacatgcccctgccaggtctcggttcccggatcggggaaagatg 2088

Qy        901 ccaccctcctgatacatgaagcacacctcctggaagatgtgtttggaaggaggaagcaggtggaa 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 ccaccctcctgatacatgaagcacacctcctggaagatgtgtttggaaggaggaagcaggtggaaa 2148

Qy        961 agacacacagcacaacgtcccaagccatcaagcctgtggggatg 1001
              |||||||||||||||||||||||||||||||||||||||||||
Db       2149 agacacagcacaacgtcccaagccatcaagcctgtggggatg 2189
```

```
RESULT  3
AAA52810
ID   AAA52810 standard; cDNA; 2478 BP.
XX
AC   AAA52810;
XX
DT   20-SEP-2000 (first entry)
XX
DE   Human sulphatase G cDNA.
XX
KW   Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH
FT   Key              Location/Qualifiers
FT   CDS              1..2478
FT                    /*tag= a
FT                    /partial
FT                    /product= "hSG"
XX
PN   WO200034327-A1.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999;  99WO-AU01092.
XX
PR   09-DEC-1998;  98AU-0007624.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ, Litjens T, Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   P-PSDB; AAY99850.
XX
PT   Novel isolated DNA sequence which encodes human sulfatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency -
XX
PS   Claim 2; Page 29-30; 33pp; English.
XX
CC   The present sequence encodes human sulphatase G (hSG). hSG is
CC   not a member of the well-characterised CTPSR sulphatase family.
CC   It belongs to a family showing sequence similarity to a sulphatase
CC   from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC   contains 23 exons and is located at chromosome 17p11.2. The present
CC   sequence is clone lambda29.1 from a human testes cDNA library. It was
CC   isolated using human EST sequences with sequence similarity to the
CC   non-CTPSR family as a probe to screen the library. The cDNA insert was
CC   subcloned and the DNA sequence of both strands was determined. The
CC   sequence may be used to treat a patient suffering from hSG deficiency
```

```
CC    by replacing, repairing, or compensating for a DNA sequence within that
CC    patient's genome.
XX
SQ    Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;

Query Match         98.3%;  Score 984.2;  DB 21;  Length 2478;
Best Local Similarity  99.6%;  Pred. No. 2.4e-265;
Matches 997;  Conservative 0;  Mismatches 3;  Indels 1;  Gaps 1;

Qy       1 tcgcagcccacaagacactcaactcatccaccggacatcttccccctgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 tcgcagcccacaagacactcaactcatccaccggacatcttccccctgct 1199

Qy      61 caccagttccgctgtaagaagaggaggcgcccaccctcagtgtgtccatggttcaggtga 120
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 caccagttccgctgtaagaagaggaggcccaccctcagtgtgtccatggttcaggtga 1259

Qy     121 atgcctcctcaagtaccagctccgtccgtccagagaggaggatgccattattac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 atgcctcctcaagtaccagctccgtccgtccagagaggaggatgccattattac 1319

Qy     181 ttgcaatccctgagggaattcatagttgaggcgctgcagcttccaactttccagagcgct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1320 ttgcaatcctgagggaattcaatagttgaggcgctgcagcttccaactttccagagcgt 1379

Qy     241 gcaggagtacagaggagagtgcgcagagacgcccagccccagagcaggaagaaagtcagta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 gcaggagtacagaggagtgcgcgcagagacgcccagccccagagcaggaaagaagtcagta 1439

Qy     301 cccagaaatcatcttccttggaacagggtctgccatcccgatgaagattccagaggcgct 360
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 cccagaaatcatcttccttggaacagggtctgccatcccgatgaagattccagaggcgt 1499

Qy     361 tgctgtgtttgtcaacataagccacctgtctctgctactggttgggttgaaggtgaggac 420
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 tactgtgttgtctcccacct-gcacgcagatcacaacaacagtgccaagtatctctgc 1559

Qy     421 atttggccagctcgtgtcgtcattacggaagacaggtctctgggcaccctggc 480
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 gtttgggcagctcgtgtcgtcattacggaagacaggtctctgggcaccctggc 1619

Qy     481 tgctgtgtttgtgtctcccacctggcacgcagatcaccacacaggtgccaagtatctgc 540
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1620 tactgtgttgtctcccacct-gcacgcagatcacaacaacagtgccaagtatctctgc 1678

Qy     541 tgcagagaagcggcctggcatcttcggaaagccgaactcaccctttgtggttg 600
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1679 tgcagagaacgcggccttggtcatcttggaaagccgaactcaccctttgtggttg 1738

Qy     601 cccccaacacagctcaaagcctggctccagcagtaccacaacaagtccaagtgaggtcctgc 660
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1739 ccccacacagctcaaaggcctggctccagcagtaccacaacaagtccaagtgaggtcctgc 1798

Qy     661 accacatcagtatgatcctgcaatgcctcaggaaggaggctgaatctccagtcctg 720
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1799 accacatcagtatgattcctgccaaatgcctcaggaagggctgagatctccagtcctg 1858

Qy     721 cagtggaaagattgatcagttcgctgttgcgaacatgtgattttggagagttcagacct 780
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1859 cagtggaaagattgatcagttcccctgttgcgaacatgtgattttggagagttcagacct 1918

Qy     781 gtctgtgtgcggcactgcaagcatgcgtttggctgtgctggtgtgtcaacctggctgga 840
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1919 gtctggtgcgcactgcaagcatgcgtttggctgtgctggtgtgtccacctggctgga 1978

Qy     841 aagtggtctattccgggacaacatcctcggaggctgccgggatgaagagaagtcgaaa 900
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1979 aagtggtctattccgggacaacatcctcgaggctctcgtcgtctgatgggaaagatg 2038

Qy     901 ccaccctcctgatacatgaagccacccctgaagagaagaagatggtttggaagagaagcagtcgaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    2039 ccaccctcctgatacatgaagccaccctgaagagaagaagatggtttggaagagaagcagtgaa 2098

Qy     961 agacacacagcacaacgtcccaagccatcaagcgtgggatg 1001
           |||||||||||||||||||||||||||||||||||||||||
Db    2099 agacacacagcacaacgtcccaagccatcaagcgtgggatg 2139

RESULT 4
AAA60207

ID   AAA60207 standard; DNA; 26664 BP.
XX
AC   AAA60207;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 genomic sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             910..26039
FT                   /number=1
FT                   /product= "HPC2"
FT                   /tag= a
FT                   /note= "this sequence contains introns"
FT                   /transl_except= (pos:23892..23895,aa:Glu)
FT   exon            910..1154
FT                   /tag= b
FT                   /number=1
FT   exon            1736..1786
FT                   /tag= c
FT                   /number=2
FT   exon            1925..1995
FT                   /tag= d
FT                   /number=3
FT   exon            3025..3089
FT                   /tag= e
FT                   /number=4
FT   exon            4361..4418
FT                   /tag= f
FT                   /number=5
FT   exon            5582..5650
FT                   /tag= g
FT                   /number=6
FT   exon            7075..7194
FT                   /tag= h
FT                   /number=7
FT   exon            8186..8244
FT                   /tag= i
FT                   /number=8
FT   exon            12878..12936
FT                   /tag= j
FT                   /number=9
FT   exon            13032..13104
FT                   /tag= k
FT                   /number=10
FT   exon            13756..13868
FT                   /tag= l
FT                   /number=11
FT   exon            15283..15378
FT                   /tag= m
FT                   /number=12
FT   exon            16278..16416
FT                   /tag= n
FT                   /number=13
FT   exon            16498..16583
FT                   /tag= o
FT                   /number=14
FT   exon            18583..18701
FT                   /tag= p
FT                   /number=15
FT   exon            20349..20445
```

```
FT                  /*tag= q
FT   exon           22172..22310
FT                  /*tag= r
FT   exon           22879..22917
FT                  /*tag= s
FT   exon           23045..23154
FT                  /*tag= t
FT   exon           23795..23895
FT                  /number= 19
FT                  /*tag= u
FT   exon           23973..24093
FT                  /number= 20
FT                  /*tag= v
FT   exon           24354..24432
FT                  /number= 21
FT                  /*tag= w
FT   exon           25026..25170
FT                  /number= 22
FT                  /*tag= x
FT   exon           25812..26036
FT                  /number= 23
FT                  /*tag= y
FT   polyA_signal   26447..26452
FT                  /number= 24
FT                  /*tag= z
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
PR   06-NOV-1998;  98US-0107468.
PN   PN
PF   05-NOV-1999;  99WO-US26055.
XX
PD   18-MAY-2000.
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  .
XX
PS   Claim 3; Page 108-122; 157pp; English.
XX
CC   The present sequence is the genomic sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;
```

```
Query Match             13.6%;  Score 136.6; DB 21; Length 26664;
Best Local Similarity   73.9%;  Pred. No. 6.2e-28;
Matches   224;  Conservative    1; Mismatches    0; Indels   78; Gaps   1;

Qy        668 cagtatgatcctgccaaatgcctcaggaaggggctggatctccagtcctgcagtgga 727
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23792 cagtatgatcctgccaaatgcctcaggaaggggctgagatctccagtcctgcagtgga 23851

Qy        728 aagattgatcaagttcgctgttgcgaacatgtgatttggaag------------------ 768
              |||||||||||||||||||||||||:|||||||||||||||||
Db      23852 aagattgatcaagttcgctgttgcgaacrtgtgatttggaagaggtaaggggcacagccgc 23911
```

```
Qy        769 ---------------------------------------------------------a 769
                                                                         |
Db      23912 aggcatcatgtggggcgagtgggggagcagaagctgcagaagcacctccagcccacccttca 23971

Qy        770 gttcagacctgtctggtcgggcactgcaagcatgcgcttggctggtggtgcgctggtgcacac 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23972 gttcagacctgtctggtcgggcactgcaagcatgcgcttggctggtggtgcgctggtgcacac 24031

Qy        830 ctctggctggaaagtgtctattccgggggaccaccatgccctgcggaggctctggtccggat 889
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24032 ctctggctggaaagtgtctattccgggggaccaccatgccctgcggaggctctggtccggat 24091

Qy        890 ggg 892
              |||
Db      24092 ggg 24094
```

```
RESULT  5
AAC34173
ID   AAC34173 standard; DNA; 1617 BP.
XX
AC   AAC34173;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 5716.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   28-APR-1999;  99US-0130891.
PR   30-APR-1999;  99US-0131449.
PR   04-MAY-1999;  99US-0132048.
PR   05-MAY-1999;  99US-0132407.
PR   06-MAY-1999;  99US-0132484.
PR   06-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   07-MAY-1999;  99US-0132487.
PR   11-MAY-1999;  99US-0132863.
PR   14-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
```

```
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  16-JUN-1999;  99US-0139453.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144321.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  21-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145089.
PR  23-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  26-JUL-1999;  99US-0145224.
PR  27-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  28-JUL-1999;  99US-0145919.
PR  02-AUG-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146366.
PR  02-AUG-1999;  99US-0146388.
PR  03-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.

PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  30-AUG-1999;  99US-0151080.
PR  31-AUG-1999;  99US-0151303.
PR  01-SEP-1999;  99US-0151438.
PR  07-SEP-1999;  99US-0151930.
PR  10-SEP-1999;  99US-0153363.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0153070.
PR  16-SEP-1999;  99US-0154018.
PR  20-SEP-1999;  99US-0154039.
PR  22-SEP-1999;  99US-0154779.
PR  23-SEP-1999;  99US-0155139.
PR  24-SEP-1999;  99US-0155486.
PR  28-SEP-1999;  99US-0155659.
PR  29-SEP-1999;  99US-0155596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  14-OCT-1999;  99US-0159929.
PR  14-OCT-1999;  99US-0159930.
PR  14-OCT-1999;  99US-0159331.
PR  18-OCT-1999;  99US-0159584.
PR  18-OCT-1999;  99US-0159638.
PR  21-OCT-1999;  99US-0159637.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  21-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  22-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  25-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  26-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  28-OCT-1999;  99US-0161993.
```

```
PR   29-OCT-1999;    99US-0162142.

     Query Match             7.0%;    Score 70;  DB 21;  Length 1617;
     Best Local Similarity  63.9%;    Pred. No. 8.1e-10;
     Matches  106;  Conservative    0;  Mismatches   60;  Indels   0;  Gaps   0;

Qy      832 ctggctggaaagtggtctattccgggagacccatctcggtccggatgg 891
            | || |||| |||||| ||| |||  ||| || || ||||||||| |||
Db     1093 cggggtggaaatggtttaccgggagacacaaggccttgtcctgagatggttgaagctt 1152

Qy      892 ggaaagatgccaccctcctcgataccatgaagccacctcggaagatggtttggaaggaag 951
            |||| ||| |||||||||||
Db     1153 caaaaggagcaacagttcttatacatgaggcgacgtttgaggatgcgttggtagaagaag 1212

Qy      952 cagtcggaaaagacaacagccaagccatcagcgtggg 997
            | |||| ||| || ||||||  ||||| ||||||| ||
Db     1213 ctgtggctaaaaccatagcacaactaagaagcaatcaaagtgg 1258

RESULT   6
AC  AAC49965
ID  AAC49965 standard; DNA; 2853 BP.
XX
AC  AAC49965;
XX
DT  18-OCT-2000  (first entry)
XX
DE  Arabidopsis thaliana DNA fragment SEQ ID NO: 63099.
XX
KW  Hybridisation assay; genetic mapping; gene expression control;
KW  protein identification; signal transduction pathway;
KW  metabolic pathway; promoter; termination sequence; ss.
XX
OS  Arabidopsis thaliana.
XX
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000;  2000EP-0301439.
XX
```

PR			PR		
PR	05-MAR-1999;	99US-0121825.	PR	25-MAY-1999;	99US-0136021.
PR	05-MAR-1999;	99US-0123180.	PR	27-MAY-1999;	99US-0136392.
PR	09-MAR-1999;	99US-0123548.	PR	28-MAY-1999;	99US-0136782.
PR	23-MAR-1999;	99US-0125788.	PR	01-JUN-1999;	99US-0137222.
PR	25-MAR-1999;	99US-0126264.	PR	03-JUN-1999;	99US-0137528.
PR	29-MAR-1999;	99US-0126785.	PR	04-JUN-1999;	99US-0137502.
PR	01-APR-1999;	99US-0127462.	PR	07-JUN-1999;	99US-0137724.
PR	06-APR-1999;	99US-0128234.	PR	08-JUN-1999;	99US-0138094.
PR	08-APR-1999;	99US-0128714.	PR	10-JUN-1999;	99US-0138540.
PR	16-APR-1999;	99US-0129845.	PR	10-JUN-1999;	99US-0138847.
PR	19-APR-1999;	99US-0130077.	PR	14-JUN-1999;	99US-0139119.
PR	21-APR-1999;	99US-0130049.	PR	16-JUN-1999;	99US-0139452.
PR	23-APR-1999;	99US-0130510.	PR	16-JUN-1999;	99US-0139453.
PR	28-APR-1999;	99US-0130891.	PR	17-JUN-1999;	99US-0139492.
PR	30-APR-1999;	99US-0131149.	PR	18-JUN-1999;	99US-0139454.
PR	30-APR-1999;	99US-0131248.	PR	18-JUN-1999;	99US-0139455.
PR	04-MAY-1999;	99US-0132407.	PR	18-JUN-1999;	99US-0139456.
PR	05-MAY-1999;	99US-0132484.	PR	18-JUN-1999;	99US-0139457.
PR	06-MAY-1999;	99US-0132485.	PR	18-JUN-1999;	99US-0139458.
PR	06-MAY-1999;	99US-0132486.	PR	18-JUN-1999;	99US-0139459.
PR	07-MAY-1999;	99US-0132487.	PR	18-JUN-1999;	99US-0139460.
PR	11-MAY-1999;	99US-0132863.	PR	18-JUN-1999;	99US-0139461.
PR	14-MAY-1999;	99US-0134256.	PR	18-JUN-1999;	99US-0139462.
PR	14-MAY-1999;	99US-0134218.	PR	18-JUN-1999;	99US-0139463.
PR	14-MAY-1999;	99US-0134219.	PR	18-JUN-1999;	99US-0139750.
PR	14-MAY-1999;	99US-0134221.	PR	21-JUN-1999;	99US-0139763.
PR	18-MAY-1999;	99US-0134370.	PR	22-JUN-1999;	99US-0139817.
PR	18-MAY-1999;	99US-0134768.	PR	23-JUN-1999;	99US-0139899.
PR	19-MAY-1999;	99US-0134941.	PR	23-JUN-1999;	99US-0140354.
PR	20-MAY-1999;	99US-0135124.	PR	24-JUN-1999;	99US-0140353.
PR	21-MAY-1999;	99US-0135333.	PR	28-JUN-1999;	99US-0140695.
PR	24-MAY-1999;	99US-0135629.	PR	29-JUN-1999;	99US-0140823.
			PR	29-JUN-1999;	99US-0140991.
			PR	30-JUN-1999;	99US-0141287.
			PR	01-JUL-1999;	99US-0141842.
			PR	01-JUL-1999;	99US-0142154.
			PR	02-JUL-1999;	99US-0142055.
			PR	06-JUL-1999;	99US-0142390.
			PR	08-JUL-1999;	99US-0142803.
			PR	09-JUL-1999;	99US-0142920.
			PR	12-JUL-1999;	99US-0142977.
			PR	13-JUL-1999;	99US-0143542.
			PR	14-JUL-1999;	99US-0143624.
			PR	15-JUL-1999;	99US-0144005.
			PR	16-JUL-1999;	99US-0144085.
			PR	16-JUL-1999;	99US-0144086.
			PR	19-JUL-1999;	99US-0144325.
			PR	19-JUL-1999;	99US-0144331.
			PR	19-JUL-1999;	99US-0144332.
			PR	19-JUL-1999;	99US-0144333.
			PR	19-JUL-1999;	99US-0144334.
			PR	19-JUL-1999;	99US-0144335.
			PR	20-JUL-1999;	99US-0144352.
			PR	20-JUL-1999;	99US-0144632.
			PR	20-JUL-1999;	99US-0144814.
			PR	20-JUL-1999;	99US-0144884.
			PR	21-JUL-1999;	99US-0145086.
			PR	21-JUL-1999;	99US-0145088.
			PR	22-JUL-1999;	99US-0145085.
			PR	22-JUL-1999;	99US-0145087.
			PR	22-JUL-1999;	99US-0145089.
			PR	22-JUL-1999;	99US-0145192.
			PR	22-JUL-1999;	99US-0145145.
			PR	23-JUL-1999;	99US-0145218.
			PR	23-JUL-1999;	99US-0145224.
			PR	26-JUL-1999;	99US-0145276.
			PR	27-JUL-1999;	99US-0145913.
			PR	27-JUL-1999;	99US-0145918.
			PR	27-JUL-1999;	99US-0145919.
			PR	28-JUL-1999;	99US-0145951.
			PR	02-AUG-1999;	99US-0146388.
			PR	02-AUG-1999;	99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   7.0%;   Score 70;   DB 21;   Length 2853;
Best Local Similarity   63.9%;   Pred. No. 1e-09;
Matches   106;   Conservative   0;   Mismatches   60;   Indels   0;   Gaps   0;

Qy   832 ctggctgggaaagtgtctattccgggggaccaccatgccctcgcaggctctcgtccgatgg   891
           ||||| |||| |||||||| || | ||||| ||| | ||||| | |||||
Db   2333 cggggtggaaaatgttactccgggagacacacaagccttgcctgagatggttgaagctt   2392

Qy   892 ggaaagatgccaccctcctggatacatgaagccaccctggaagatggtttggaagaggaag   951
           |||| | || | || | | | | | ||||| | ||| | | |||| |||
Db   2393 caaaagagagcaaacagttcttatacatgaggcgacgtttgaggatgccgttggtagaagaag   2452

Qy   952 cagtggaaaagacacacagccaacgccatcagcgtggg   997
           ||||| || | | || ||| ||||| || ||
Db   2453 ctgtggctaaaaccatagcacacaactaaagaagcaatcaaagtgg   2498

RESULT   7
AAC38171
ID   AAC38171 standard; DNA; 464 BP.
XX
AC   AAC38171;
XX
DT   17-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 20026.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137228.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 21-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142290.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145008.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0151066.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

```
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
XX
Qy   Query Match     4.2%;   Score 41.6;   DB 21;   Length 464;
     Best Local Similarity   54.6%;   Pred. No. 0.042;
Matches   83;   Conservative   0;   Mismatches   69;   Indels   0;   Gaps   0;
XX
Qy  281 gcagagaaagaaagtcagtaccaggaaatcatcttccttggaacaaggggtctgccatccg 340
            |  ||| ||| | || ||||  |||| |  |  |||||| |  ||
Db   89 gaaaagattatacgagatgacatggagattgttattccaggccactggttcgtctcagct 148
XX
Qy  341 atgaaagattcagtgccacacttgtcaaacatacaagccacgcctgccactctcgcta   400
            ||||| | |||| | |  ||||| ||  |||| | ||||| || |
Db  149 tcaaaataccgaaacgttagtgcaattttatttgactttgtctctagaggtagtcttctc 208
XX
Qy  401 ctggactgtggtcgaggcacattttgggccagct 432
            |  ||||| |||||||| |||  ||||||| ||
Db  209 ctagattgtggtgaaggaaacttgggccagct 240
```

RESULT 8

AAC04378

ID   AAC04378 standard; cDNA; 521 BP.
XX
AC   AAC04378;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 8453.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000;   2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 8453; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 521 BP; 122 A; 187 C; 87 G; 123 T; 2 other;

```
Qy   Query Match     3.8%;   Score 37.6;   DB 21;   Length 521;
     Best Local Similarity   57.8%;   Pred. No. 0.58;
Matches   67;   Conservative   0;   Mismatches   49;   Indels   0;   Gaps   0;
XX
Qy   30 acctcatccaccccggacaatcttcccctgctcaccagttccgcgtgtaagaaggaggcc  89
           ||| ||||||||||| |||||| |||| |||||||||||| ||||| |||||
Db  187 accatatccaacctggaccgctggtctcctcctcaaccttttggccaggattgttccacc 246
XX
Qy   90 ccacccctcagtgttgcccatggttgaatgctcggtcaagtactaccagctccgt  145
           ||| |||||||| | ||| ||||||| || | |||||| ||| |||||
Db  247 acctcctcctccacccctatggtccagggaggaaatcccacactcctcctcccgccaccct 302
```

RESULT 9

AAC89621

ID   AAC89621 standard; DNA; 2517 BP.
XX
AC   AAC89621;
XX
DT   08-MAR-2001 (first entry)
XX
DE   S. cerevisiae YKR079C gene.
XX
KW   Yeast; germination; proliferation; essential gene; antifungal agent;
KW   insecticide; herbicide; anti-proliferation drug; cancer; psoriasis;
KW   restenosis; YKR081C; YFR003C; YGR277C; YGR278W; YKR071C; YKR079C;
KW   YKR083C; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200071161-A1.
XX
PD   30-NOV-2000.
XX
PF   12-MAY-2000;   2000WO-US13017.
XX
PR   21-MAY-1999;   99US-0315794.
PR   02-SEP-1999;   99US-0389441.
XX
PA   (ROSE-) ROSETTA INPHARMATICS INC.
XX
PI   Roberts CJ;
XX
DR   WPI; 2001-025092/03.
DR   P-PSDB; AAB49964.
XX
PT   Identifying antifungal compounds which target yeast essential genes
PT   comprises use of novel Saccharomyces cerevisiae essential genes
PT   YFR003C, YGR277C, YGR278W, YKR071C, YKR079C or YKR081C -
XX
PS   Example 5; Fig 29; 127pp; English.
XX
CC   The present invention provides methods of identifying antifungal agents
CC   using the coding and protein sequences of several yeast genes. These are
CC   essential for the germination and proliferation of Saccharomyces
CC   cerevisiae, and include YKR081C, YFR003C, YGR277C, YGR278W, YKR071C,
CC   YKR079C and YKR083C. The sequences can also be used to identify compounds
CC   for use as herbicides, insecticides and anti-proliferation drugs which
CC   can be used in the treatment of cancer, psoriasis and restenosis. This is
CC   because they can be used to identify plant, insect and human homologues
CC   of the yeast genes.
XX
SQ   Sequence 2517 BP; 881 A; 417 C; 496 G; 723 T; 0 other;

```
Qy   Query Match     3.7%;   Score 37;   DB 22;   Length 2517;
     Best Local Similarity   57.3%;   Pred. No. 1.6;
Matches   50;   Conservative   0;   Mismatches   1.6;   Indels   0;   Gaps   0;
XX
Qy  883 tccggatgggaagaatgccaccctcctgatacatgaagcacccctgggaagatggtttgg  942
           || ||| ||| |||||||| ||| | |||||| || |||||| | ||||| |||
Db 2126 tcgaaataggctataattcagatctattcacgagaagctacactagaaaatcagctac 2185
```

Qy   943 aagaggagagcagtggaaagacacaagcatccaagccatcaagagtgggga 999
                ||||| |||| ||| ||||| |||| || |||| |||| ||| ||| ||
Db   2186 tggaggatgccgtgaagagaaaaacactgcactattaatgaagcaatcggtgtttcga 2242

XX
DT   29-JUN-2000 (first entry)
XX
AC   AAA10594;
XX
ID   AAA10594 standard; DNA; 10732 BP.
RESULT   10
Qy

DE   Gene encoding a subunit of cellulose synthase.
XX
KW   Cellulose synthase; cellulose production; increase yield; ds.
XX
OS   Vigna angularis.
XX
PN   JP200060568-A.
XX
PD   29-FEB-2000.
XX
PF   26-AUG-1998;   98JP-0239998.
XX
PR   26-AUG-1998;   98JP-0239998.
XX
DR   WPI; 2000-342371/30.
DR   P-PSDB; AAY85179.
XX
CC   This sequence represents a gene encoding a subunit of the cellulose
CC   synthase complex of Vigna angularis. The invention relates to subunits of
CC   cellulose synthetic equipment, that can be used to increase the amount of
CC   cellulose synthesised by a plant. The proteins and genes encoding them
CC   can also be used to improve the properties of the cellulose being
CC   produced by a plant.
XX
PT   A gene encoding a cellulose synthetic equipment - for the improvement
PT   in the amount of cellulose synthesised in a plant body
XX
PS   Claim 2; Page 14-21; 32pp; Japanese.
XX
SQ   Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match   3.6%; Score 36.4; DB 21; Length 10732;
Best Local Similarity 15.6%; Pred. No. 4.4;
Matches 139; Conservative 325; Mismatches 413; Indels 15; Gaps 3;

Qy   117 gtgaatgcctcctcaagtaccacagtctccgtcccaggaggggatgccatta 176
                || |||| | | | ||| || | | ||::
Db   9042 gngnargvatyrgargvaaahsasgyaaysgycysmtgngnarghvaaargggngngnarg 9101

Qy   177 ttacttgcaatcctgaggaattcatagttgaggcgctgcagcttcccaactccagcaga 236
                | |::| || ||| |::| | || |||::| |::| | || || | |
Db   9102 hsasaavamtysaasrysghgysgargmtvaggngnsrgthrsysargaaysaagvaasa 9161

Qy   237 gcgtcgcaggagagtacagcgggcgacgccagcaggcccacagaaaagaaagtc 296
                : | ||:: || ||| ::||| ||| | | ||| | | |::| :::|
Db   9162 snsrhsasnyssrhaaargggasngnasargnghrgnaaggashsgyaamtsrargg 9221

Qy   297 agtaccacagaaatcatccttcccttggaacagggtctgtcccatcccgatgaagattcgaaatg 356
                :::::|| | ::|: || ||||::|| | ||||:: | |:| ||| |||::
Db   9222 srmtgysasasnthrsrthrtyrgvaargvagysggargasnggargghsrargargthr 9281

Qy   357 tcagtgccacacttgtcaacataagcccgacacgtctctgctactg-gactgtggtgag 415
                : :|:||: ::| | | :| :::|| |:: | ||||:| :| | ||
Db   9282 aagvasrhsysgngngsrthrysysaaysggcysgnaargvaargysargrgyraaaa 9341

Qy   416 ggcacatttgggcagctgtgccgtcattacgagaaggacagtggacagggtcctgggcacc 475
                : :| ||:||||:|: :| |:
Db   9342 aaysmtysaasngyagmtgyargassrhgaargasryssrthrasnsrvagsrsrvaast 9401

Qy   476 ctgg----------ctgctgttgttgtgtcccacctgaagcacgcgaatcacacagtgc 527
                |::: | |: | | ||| | |::: :| | | :: | :
Db   9402 hrsrrgthrrthrysasnthrthrgngncysaavaggggasnysvaaysgsrasnargysasn 9461

Qy   528 ccaagtatcttgctgcagagaacgccgcctggcatcttgggaaagccgcttcacct 587
                :::::::||| | | :|| || | | ||| | | ||:: | || | |::|
Db   9462 ggnhsrargvaamtthrargthraaasryshsgsrgngashsysgyhsssrvaghshssrasn 9521

Qy   588 ttgctggttgttgccccccaaccagctcaaagcctgtccagcagtaccacacagtgc 647
                || :| || || || || | || | | || || ||:::| || :::
Db   9522 vaasrhsghsthaasrsrasaagysrasasysthrsrcysaasgysrtraasrasrgash 9581

Qy   648 caggaggtcctgcaaccacatcagtatgatcctgccaaatgcctcaggaaagggctgag 707
                || || ||| :| | :::| ||| | ||| || || || || | |::| :::|
Db   9582 shargsrgysysysgysrcysysysysrvagyaasrasasnmtasashvagmtysaavavas 9641

Qy   708 acctctggctgaaagatggtctattccgggtggacaacatgtcgattttggaa 767
                ::||| | ||| || |||| | |::| | | || ||:| || |||| ||
Db   9642 rvagysaathrqnsrsrtraasrysaavathrqasnsrgthrgythrysgsrthrrggg 9701

Qy   768 gagttcaagacctgtcggtgcggcactgcaagcatcgcgttggctgtcgctggtgcac 827
                :| | |::|||||| :| |::|| | |||:| | | ||::|||| ||||| ::|
Db   9702 yysgrvarashssrgnthrthrghthrthrrgvavaasntgrmtsrashsaasrysysry 9761

Qy   828 acctctggctggaaagtggtctattccgggtggacaactgcccctgcgaggctctgtccgg 887
                :::||| ||||| ||||| | |::| || | ||||| ||| | |:||| ||||:::|
Db   9762 sasnthrcyssrashsysggythrvargytrgnasvavaysthrggnasnhsvathrhsys- 9820

Qy   888 atggggaaagtgccacacctcctgataacatgaagccaccctggaagatggtttggacgagag 947
                : |::| | |::| ||| | | || | ||:| | ||| |::| :: | |:|::|
Db   9821 ----gyrasashsasysvaaaarghsvaasnashashhassrsraargygysrvahsas 9875

Qy   948 gaagcagtggaaagacacacagcacaagcgtcccaagccatcaagcgtggggga 999
                :|:|| || || | ::: | || ::|::| ||:::| :| | ||
Db   9876 thrgnasnrrgncyshscyssrccyssrcyssrassrsrvavaasnrsraagygasna 9927

XX
DT   17-OCT-2000 (first entry)
XX
AC   AAC42402;
XX
ID   AAC42402 standard; DNA; 1409 BP.
RESULT   11
AAC42402/c

DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 35418.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 24-JUN-1999; 99US-0140353.
PR 28-JUN-1999; 99US-0140354.
PR 29-JUN-1999; 99US-0140695.
PR 30-JUN-1999; 99US-0140823.
PR 01-JUL-1999; 99US-0140991.
PR 02-JUL-1999; 99US-0141287.
PR 06-JUL-1999; 99US-0141842.
PR 08-JUL-1999; 99US-0142154.
PR 09-JUL-1999; 99US-0142055.
PR 12-JUL-1999; 99US-0142390.
PR 13-JUL-1999; 99US-0142803.
PR 14-JUL-1999; 99US-0142920.
PR 15-JUL-1999; 99US-0142977.
PR 16-JUL-1999; 99US-0143542.
PR 16-JUL-1999; 99US-0143624.
PR 19-JUL-1999; 99US-0144005.
PR 19-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160815.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160980.
PR   21-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   22-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   25-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
```

Query Match

Best Local Similarity   60.0%;

Matches   60; Conservative   0; Mismatches   2.5; Indels   0; Gaps   0;

```
Qy   574 agccgcttcaccctttgctggttggttgccccaacaagctcaaagcctggcctccacagcagt 633
         || | ||| | ||  |||  || |||| |  ||| || | |||| | ||| ||| ||||| |||
Db   350 AGGACCTGAACCACTTGATGGTATTGACCCTGTACCTGCACCGAGACCGCTTCCAGTAGA 291
```

```
Qy   634 accaacaacaagtgccaggaggtcctgcaccacatcagtat 673
         |||  |||||||
Db   290 ACCTGAACCAGATGACCCTGATCCTGCACCAAAACCGAAT 251
```

RESULT   12

AAA59146/c

ID   AAA59146 standard; DNA; 19053 BP.

XX

AC   AAA59146;

XX

DT   07-NOV-2000 (first entry)

XX

DE   Nucleotide sequence of the antibiotic Tel-aviv gene cluster.

XX

KW   Polyketide; antibiotic Tel-Aviv; cell wall synthesis;

KW   lipid-disaccharide-pentapeptide; gingivitis; ss.

XX

OS   Myxococcus xanthus.

XX

FH   Key   Location/Qualifiers

FH

FT   CDS   complement (2955..601)

FT   /*tag= a

FT   CDS   complement (4702..3116)

FT   /*tag= b

FT   /note= "encodes AAB07666"

FT   CDS   complement (5595..4720)

FT   /*tag= c

FT   /note= "encodes AAB07667"

FT   CDS   6290..6793

FT   /*tag= c

FT   /note= "encodes AAB07668"

FT   CDS   6870..7106

FT   /*tag= e

FT   /note= "encodes AAB07669"

FT   CDS   7119..8378

FT   /*tag= f

FT   /note= "encodes AAB07670"

FT   CDS   8404..9378

FT   /*tag= g

FT   /note= "encodes AAB07671"
```

```
FT                   /*tag= g
FT   CDS             9386..9364
FT                   /*tag= h
FT                   /note= "encodes AAB07672"
FT   CDS             9757..10878
FT                   /*tag= i
FT                   /note= "encodes AAB07673"
FT   CDS             10909..11421
FT                   /*tag= j
FT                   /note= "encodes AAB07674"
FT   CDS             11473..12897
FT                   /*tag= k
FT                   /note= "encodes AAB07675"
FT   CDS             12938..13891
FT                   /*tag= l
FT                   /note= "encodes AAB07676"
FT   CDS             13909..14898
FT                   /*tag= m
FT                   /note= "encodes AAB07677"
FT   CDS             14963..16213
FT                   /*tag= n
FT                   /note= "encodes AAB07678"
FT                   /trans1_except= (pos: 15521..15523, aa: Lys)
FT                   /note= "encodes AAB07679"
FT   CDS             16224..17009
FT                   /*tag= o
FT                   /trans1_except= (pos: 16593..16595, aa: Ser)
FT                   /note= "encodes AAB07680"
FT   CDS             17000..17767
FT                   /*tag= p
FT                   /note= "encodes AAB07681"
FT   CDS             17782..19053
FT                   /*tag= q
FT                   /note= "encodes AAB07682"
```

XX

PN   EP1026248-A2.

XX

PD   09-AUG-2000.

XX

PF   31-JAN-2000; 2000EP-0300747.

XX

PR   29-JAN-1999;   99US-0240537.

XX

PA   (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

XX

PI   Rosenberg E, Ron E, Orr E, Paltan Y;

XX

DR   WPI; 2000-500254/45.

DR   P-PSDB; AAB07666, AAB07667, AAB07668, AAB07669, AAB07670, AAB07671,

DR   AAB07672, AAB07673, AAB07674, AAB07675, AAB07676, AAB07677, AAB07678,

DR   AAB07679, AAB07680, AAB07681, AAB07682.

XX

PT   Novel DNA sequence involved in polyketide antibiotic Tel-Aviv

PT   production useful for inhibiting cell wall synthesis and in wide range

PT   of clinical applications such as treating gingivitis   -

XX

PS   Disclosure; Page 12-19; 66pp; English.

XX

CC   The specification describes a DNA sequence which partially encodes

CC   a functional portion of polypeptide component required for synthesizing

CC   the polyketide antibiotic Tel-Aviv, postmodification of antibiotic

CC   Tel-aviv, or regulation of biosynthesis of antibiotic Tel-aviv. The

CC   antibiotic Tel-aviv is a macrocyclic polyketide synthesised through

CC   the incorporation of acetate, methionine, and glycine. It inhibits cell

CC   wall synthesis by interfering with the polymerisation of the

CC   lipid-disaccharide-pentapeptide. Antibiotic Tel-aviv genes are useful

CC   in combinatorial genetics, and for encoding protein components for the

CC   synthesis, modification and regulation of antibiotic Tel-aviv.

CC   Antibiotic Tel-aviv is useful in a wide range of clinical applications

CC   such as treating gingivitis. Antibiotic Tel-aviv is also useful for

CC   generating new biological agents from its secondary metabolites. The

CC   present sequence encodes proteins involved in synthesis of antibiotic

CC   Tel-aviv.
```

```
XX
SQ   Sequence 19053 BP; 2947 A; 6220 C; 6640 G; 3246 T; 0 other;

Query Match     3.6%;  Score 35.8;  DB 21;  Length 19053;
Best Local Similarity   45.8%;  Pred. No. 8.2;
Matches  124;  Conservative   0;  Mismatches  147;  Indels   0;  Gaps   0;

Qy         28 caacctcatccaccgagacatcttccccgtgctccaccagttccgctgtaagaaggaggg 87
Db       3091 CCACCCACCTTCCAGGCGACATCTGTCCCCTGACCACCCGTCTTTTCTGGTATCCGCTTG 3032

Qy         88 cccaccctcagtcgtgccatggttcaggttgaatgcctctcaagtaccagctccgtcc 147
Db       3031 TTCCGAGCTGGGCAAACGGCGTACGGACGAAGAGACCTTCGGCCCCACGTGCAGGAGGCTTCA 2972

Qy        148 caggaggggagtggcaggagggggatgccattattacttgcaatcctgaggaattcatagttga 207
Db       2971 GCCCAGGGGGGACCGCGCATGAAAAGTGGTCAACAAGCTGTGGAGAAGCTTCCCACGTCGT 2912

Qy        208 ggcgctgcagcttcccaacttccagcagaagcgtgcaggagagagtgcgcagga 267
Db       2911 GCCGGGCAAGGTGCCCGACGTGAAGCTGCCAGGACCACATCAAGGTGCCGCTGGCCCA 2852

Qy        268 cggcccagcccagcaagaagaaagaagtcag 298
Db       2851 GGGGACTTCACGGAAGAAGAGAAAATCCTTCCG 2821
```

```
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
```

```
RESULT 13
AAV89531
ID   AAV89531 standard; cDNA; 366 BP.
XX
AC   AAV89531;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
DE   EST clone CP314.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;  98WO-US06955.
XX
PR   10-APR-1997;  97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI   Racie LA, Spaulding V, Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 247; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
```

```
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 366 BP; 100 A; 85 C; 81 G; 81 T; 0 other;

Query Match     3.5%;  Score 35.4;  DB 20;  Length 366;
Best Local Similarity   59.4%;  Pred. No. 2.1;
Matches   60;  Conservative   0;  Mismatches  41;  Indels   0;  Gaps   0;

Qy         30 acctcatccaccgagacatcttcccccgtgctccaccagttccgctgtaagaaggaggcc 89
Db        167 accatatccaccggaccgctggtccctcctcaacctttggccaggattgttccaac 226

Qy         90 ccacctcagtcgtgccatgggttcaggtgaatgcctctc 130
Db        227 acctcctccaccctatggtccaggaggagaatccacctc 267
```

```
RESULT 14
AAF26315
ID   AAF26315 standard; DNA; 1716 BP.
XX
AC   AAF26315;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Pseudomonas sp lipopeptide synthase DNA ORF00203.
XX
KW   Lipopeptide synthase; transgenic plant; antiviral; antibacterial;
KW   antifungal; surfactant; microbial depollution; emulsifier;
KW   tertiary crude oil recovery; ORF00203; ds.
XX
OS   Pseudomonas sp.
XX
PN   WO200107589-A2.
XX
PD   01-FEB-2001.
XX
PF   21-JUL-2000;  2000WO-EP07002.
XX
PR   27-JUL-1999;  99DE-1035106.
XX
PA   (TIGR-) TIGR INST GENOMIC RES.
PA   (QIAG-) QIAGEN GMBH.
PA   (GBFB-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA   (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
PI   Fraser CM, Tuemmler B, Hoheisel J, Duesterhoeft A, Hilbert H;
PI   Timmis KN, Moore E, Straetz M, Heim S, Golyshin P;
XX
DR   WPI; 2001-159715/16.
XX
PT   New DNA encoding a bacterial lipopeptide synthase, useful e.g. for
PT   producing antimicrobial agents, surfactants with increased
PT   bioavailability in microbial depollution methods, and transgenic plants
PT   .
XX
PS   Claim 3a; Page 24-25; 33pp; German.
XX
CC   This invention describes a novel DNA sequence (I) that expresses products
CC   having the biological function of lipopeptide synthase. The invention
CC   also describes (II) recombinant expression vectors containing (I); (2)
CC   prokaryotic and eukaryotic cells transformed or transfected with (I) or
CC   the vector of (2); (3) production of lipopeptide synthases by culturing
CC   cells of (2); (4) expression products (II) of (I), and synthetic proteins
CC   or peptides with the same sequences; (5) mono- or poly-clonal antibodies
CC   (Ab) specific for (II); (6) hybridoma cells that produce monoclonal Ab;
```

CC   and (7) transgenic plants that contain cells of (2). (I), and their
CC   fragments, are useful for expression of recombinant lipopeptide
CC   synthases, and as probes and primers for detection, isolation and
CC   amplification of full-length cDNA sequences. (I) are used to produce
CC   transgenic plants. Lipopeptide synthases are useful for production of
CC   antiviral, antibacterial or antifungal lipopeptides, particularly useful
CC   for treatment of plants, and for production of lipopeptide surfactants
CC   used for increasing bioavailability in microbial depolution processes
CC   (acting as emulsifiers) and in tertiary crude oil recovery. Individual
CC   subunits from different lipopeptide synthases may be combined to allow
CC   synthesis of many different biologically active substances.
XX
SQ   Sequence 1716 BP; 307 A; 504 C; 560 G; 344 T; 1 other;

Query Match   3.5%;   Score 35.2; DB 22; Length 1716;
Best Local Similarity   49.0%;   Pred. No. 4.5;
Matches   94; Conservative   0; Mismatches   98; Indels   0; Gaps   0;

Qy   30 acctcatccaccggacatcttcccctgctcaccagttccgctgtaagaaggaggc 89
            |||||||| | || ||||| | || ||| | |||| | || || ||| ||||
Db   11 acctcatgatcccctcacctctcgctgaccaaaatgggcgtcgaatcaggcgaagcc 70

Qy   90 ccaccctcagtgtgtgccatggttcaggggtgaatgccctccaagtaccagctcgtccca 149
            ||| | |||| | || | | ||||| | | ||| | || || ||| ||||| | |
Db   71 tggacctttatctgtcgtccgcaaggagtgggagcgtcagccaccgggaccggtcttttcctct 130

Qy   150 ggaggagagtggcagaggaagaatgcattattacttgcaatcctgaggaattcatagttgagg 209
            ||| | || || | | | || | || | ||| | || || | ||| | | | |
Db   131 ggggcaattggccagtgcgtcttggccgacaatgctcgcgatgctgcggcatccattgatggtg 190

Qy   210 cgctgcagcttc 221
            ||||| |||
Db   191 aaatgcaggttc 202

CC   The inventors claim a maize optimised coding sequence for Bacillus
CC   thuringiensis Bt) insecticidal protein CryIA(b). Claimed sequences
CC   are AAQ39019 and AAQ39020.
XX
SQ   Sequence 3474 BP; 749 A; 1255 C; 1023 G; 447 T; 0 other;

Query Match   3.5%;   Score 35; DB 14; Length 3474;
Best Local Similarity   46.2%;   Pred. No. 6.8;
Matches   116; Conservative   0; Mismatches   135; Indels   0; Gaps   0;

Qy   27 tcaacctcatcaccggacacatcttcccctgctcaccagttccgctgtaagaaggag 86
            |||| ||| ||| || | || || || | | ||| ||| ||| | || ||||
Db   2468 tcaagatcaagaccaggacggccacgccccgcctggggcaaacctggagttccctgggaggaga 2527

Qy   87 gccccacctcagtgtgccccatggttcaggggtgaatgcctccaagtacccagctccgtc 146
            ||| | ||||| | || || || ||||| | | | || || ||| | || ||| |
Db   2528 agccccctggttgggcgggacgccctggcccgctgaagccgccggagaagaagtggcgcgaga 2587

Qy   147 caaggaggagtggcaaaggagatgccattattacttgcaatcctgaggaattcatagttg 206
            || || || |||| ||||| | || ||| | | ||| | || ||| ||| | || |
Db   2588 agcgggaggaagctggagtgggagaccaacatcgtgtacaagaggaggccaaggaggagctgg 2647

Qy   207 aggcgctgcagcttcccaaacttccagcagagcgtgcaggagtacagggagagtcgcagg 266
            || | | || | || || || || ||| || | ||||| | | || ||| || |
Db   2648 acgccccgttcgtgaacagccagtacgaccgcctgcaggccgaccaacaatcgccatca 2707

Qy   267 acggcccagcc 277
            || | |||||
Db   2708 tccaacgccgcc 2718

RESULT 15
ID   AAQ39018 standard; DNA; 3474 BP.
XX
AC   AAQ39018;
XX
DT   13-AUG-1993 (first entry)
XX
DE   Sequence of pure maize optimized synthetic Bt CryIA(b) gene.
XX
KW   Bacillus thuringiensis; CryIA(b); insecticide; ss.
OS   Synthetic.
XX
PN   WO9307278-A.
XX
PD   15-APR-1993.
XX
PF   05-OCT-1992;   92WO-US08476.
XX
PR   04-OCT-1991;   91US-0772027.
PR   25-SEP-1992;   92US-0951715.
XX
PA   (CIBA) CIBA GEIGY AG.
XX
PI   Crossland LD, Desai NM, Evola SV, Koziel MG, Kramer VC;
PI   Launski L, Lewis KS, Merlin EJ, Rothstein SJ, Warren GW;
PI   Wright MS;
XX
DR   WPI; 1993-13467/16.
XX
PT   DNA sequences, enclosing insecticidal proteins for controlling
PT   Coleoptera and Lepidoptera pests - optimised for expression in
PT   plants, esp. maize
XX
PS   Disclosure; Pages 167-169; 286pp; English.

```
OM nucleic - nucleic search, using sw model

                 Copyright (c) 1993 - 2000 Compugen Ltd.

                 GenCore version 4.5

Run on:          October 28, 2001, 21:03:52 ; Search time 50.71 Seconds
                                              (without alignments)
                                              3736.944 Million cell updates/sec

Title:           SEQLINS_COPY_1140_2140
Perfect score:   1001
Sequence:        1 tcgcgccacaagattcaaa.................aagccatcaagcgtggggatg 1001
Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched:        324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters:        649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

                 Pred. No. is the number of results predicted by chance to have a
                 score greater than or equal to the score of the result being printed,
                 and is derived by analysis of the total score distribution.

                              SUMMARIES

                   %
Result          Query
No.    Score    Match Length DB ID                Description
-----  -----    -----  -----  --  --               -----------
1       37      3.7   2517   4   US-09-315-794-51   Sequence 51, Appli
2       37      3.7   2517   4   US-09-389-341-51   Sequence 51, Appli
3       35      3.5   3546   1   US-07-951-715A-14  Sequence 14, Appl
4       35      3.5   3546   2   US-08-459-448A-14  Sequence 14, Appl
5       35      3.5   3546   3   US-08-459-595A-14  Sequence 14, Appl
6       35      3.5   3546   3   US-08-459-504B-14  Sequence 14, Appl
7       35      3.5   3546   4   US-08-459-444-14   Sequence 14, Appl
8       35      3.5   8854   3   US-09-053-549-1    Sequence 1, Appl
9       33.4    3.3   3468   3   US-07-951-715A-2   Sequence 2, Appl
10      33.4    3.3   3468   1   US-07-951-715A-4   Sequence 4, Appl
11      33.4    3.3   3468   2   US-08-459-448A-2   Sequence 2, Appl
12      33.4    3.3   3468   2   US-08-459-448A-4   Sequence 4, Appl
13      33.4    3.3   3468   3   US-08-459-595A-2   Sequence 2, Appl
14      33.4    3.3   3468   3   US-08-459-595A-4   Sequence 4, Appl
15      33.4    3.3   3468   3   US-08-459-504B-2   Sequence 2, Appl
16      33.4    3.3   3468   3   US-08-459-504B-4   Sequence 4, Appl
17      33.4    3.3   3468   4   US-08-459-444-2    Sequence 2, Appl
18      33.4    3.3   3468   4   US-08-459-444-4    Sequence 4, Appl
19      33.4    3.3   3468   3   US-09-053-549-3    Sequence 3, Appl
20      33.4    3.3   3468   6   US-09-053-549-5    Sequence 5, Appl
21      33.4    3.3   3840   6   5510474-1          Patent No. 5510474
22      33      3.3   3483   3   US-08-800-599A-23  Sequence 23, Appl
23      32.8    3.3   1872   3   US-08-422-108-2    Sequence 2, Appl
24      32.8    3.3   2385   2   US-09-146-283-3    Sequence 3, Appl
25      32.8    3.3   2385   3   US-08-579-823A-3   Sequence 3, Appl
26      32.8    3.3   2385   4   US-09-344-195-3    Sequence 3, Appli
27      32.8    3.3   3768   2   US-08-625-101-1    Sequence 1, Appli
28      32.8    3.3   3768   2   US-08-356-786-1    Sequence 1, Appli
29      32.8    3.3   4473   2   US-09-048-804-1    Sequence 1, Appli
30      32.8    3.3   4530   1   US-08-229-515A-9   Sequence 9, Appli
31      32.8    3.3   4530   1   US-08-645-865-9    Sequence 9, Appli
32      32.6    3.3   1848   1   US-08-333-802-1    Sequence 1, Appli
33      32.6    3.3   2502   1   US-08-073-384C-7   Sequence 7, Appli
34      32.6    3.3   2502   1   US-08-254-359A-7   Sequence 7, Appli
35      32.6    3.3   2502   2   US-08-483-043-7    Sequence 7, Appli
36      32.6    3.3   2502   1   US-08-481-238-7    Sequence 7, Appli
37      32.6    3.3   2502   2   US-08-471-066B-7   Sequence 7, Appli
38      32.6    3.3   2502   2   US-08-484-956-7    Sequence 7, Appli
39      32.6    3.3   2502   1   US-08-757-653-7    Sequence 7, Appli
40      32.6    3.3   2502   1   US-08-599-491-7    Sequence 7, Appli
41      32.6    3.3   2502   2   US-08-756-386-7    Sequence 7, Appli
42      32.6    3.3   2502   2   US-08-823-516-7    Sequence 7, Appli
43      32.6    3.3   2502   3   US-08-682-853A-7   Sequence 7, Appli
44      32.6    3.3   2502   3   US-08-759-038-7    Sequence 7, Appli
45      32.6    3.3   2502   3   US-08-758-314-7    Sequence 7, Appli
```

```
                                                               ALIGNMENTS

RESULT 1
US-09-315-794-51
Sequence 51, Application US/09315794
Patent No. 6197517
GENERAL INFORMATION:
APPLICANT: Roberts, Christopher J.
TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
TITLE OF INVENTION: DRUGS
FILE REFERENCE: 9301-053
CURRENT APPLICATION NUMBER: US/09/315,794
CURRENT FILING DATE: 1999-05-21
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 51
LENGTH: 2517
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-09-315-794-51

Query Match           3.7%;  Score 37;  DB 4;  Length 2517;
Best Local Similarity   57.3%;  Pred. No. 0.23;
Matches   67;  Conservative   0;  Mismatches   50;  Indels   0;  Gaps   0;

Qy    883 tccggatgggagaaagatgtgccacctcctcgatacatgaagccaccctgaaagatggtttgg 942
          || ||||| |||| ||| ||||| ||||||||||||| | |||| ||||||||| ||| |||
Db   2126 tcgaaataggctataatccagatcttattcacgaagctacactagaaaatcagctac 2185

Qy    943 aagaggaagcagtggaaaagacacacagacaaacgtccaagccatcagcgtggggga 999
          |||| || || ||||||| |||| || ||||||| |||||| ||| ||| || ||
Db   2186 tggaggatgccgtgaagagaaaaaacactgcactattaatgaagcaatcggtgtcga 2242
```

```
RESULT 2
US-09-389-341-51
Sequence 51, Application US/09389341
Patent No. 6200803
GENERAL INFORMATION:
APPLICANT: Roberts, Christopher J.
TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
TITLE OF INVENTION: DRUGS
FILE REFERENCE: 9301-057
CURRENT APPLICATION NUMBER: US/09/389,341
CURRENT FILING DATE: 1999-09-02
EARLIER APPLICATION NUMBER: 09/315,794
EARLIER FILING DATE: 1999-05-21
NUMBER OF SEQ ID NOS: 72
SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 51
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-389-341-51

Query Match   3.7%;   Score 37;   DB 4;   Length 2517;
Best Local Similarity   57.3%;   Pred. No. 0.23;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy   883 tccggatggggaaagatgcacacctcctgatacatgaagccaccctggaagatggtttgg 942
          |||||| ||| ||||| || || ||| ||| ||||| |||||||| ||||||| ||||
Db 2126 tcgaaataggctataattcagatctattaattcacgaagctacactagaaatcagctac 2185

Qy   943 aagaggagaagcaagtggaaaagacacacaacacgtcccaagccatcagcgcgtgggga 999
          ||||| | || |||| || |||||| || || ||| || |||||| || || || ||
Db 2186 tggaggatgccgtgaagaagaaaaactgcactattaatgaagcaatccggtgtgttcga 2242

RESULT 3
US-07-951-715A-14
; Sequence 14, Application US/07951715A
; Patent No. 5625136
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/951,715A
; FILING DATE: 25-SEP-1992
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3546 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3543
; OTHER INFORMATION: /product= "Full-length, hybrid,
; OTHER INFORMATION: maize optimized heat stable cryIA(b)"
; OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pCIB5513."
US-07-951-715A-14

Query Match   3.5%;   Score 35;   DB 1;   Length 3546;
Best Local Similarity   46.2%;   Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy   27 tcaacctcatccaccggacatcttcccccgttcaccagttccgctgtaagaaggagg 86
          ||| || || || || ||||| | || || | ||| |||| ||||||| |||
Db 2546 TCAAGATCAAGACCCAGGACGGCCCACGCCCCGCTCGGGCAATCTAGAGTTCCTGGAGGAGA 2605

Qy   87 gcccacccttcaagtgtgcccatggttcaggtgaatgcctcctcaagtaccagctccgtc 146
          ||||| |||||| || ||| || ||||| |||||| || || ||||| ||||||| |
Db 2606 AGCCCCTGGGTGGGGCGAGGCCCTGGCCCGGTGAAGCGCCGAAGAAGCTGGCGCCGGACA 2665

Qy   147 ccaggaggaggtggcagaggaggatgccattattacttgcaatcctgaggaattcaagttg 206
          ||||||||||| || |||||| ||| || || ||| | |||| | || || ||||| ||
Db 2666 AGCGCCGAGAAGCTGGAGTGGGACAACCATCGTGTACAAGGAGGAGGCCAAGGAGGGTGG 2725

Qy   207 aggccgctgcagcttcccaacttccagcagaggaggagagacagaggaggagtgcgcagg 266
          || || || ||| || || || |||||| || || || | ||||| || |||| ||
Db 2726 ACGCCCTGTTCGTGAACACGCCAGTACGACCGCCTGCAGGCGCGACCACCAACATCGCCATGA 2785

Qy   267 acggcccagcc 277
          | ||||| |||
Db 2786 TCCACGCCGCC 2796

RESULT 4
US-08-459-448A-14
; Sequence 14, Application US/0845944BA
; Patent No. 5859336
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5859336artis Corporation
; STREET: Patent & Trademark Dept., 520 White Plains
; STREET: Rd., POB 2005
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:

RESULT 5
US-08-459-595A-14
; Sequence 14, Application US/08459595A
; Patent No. 601810A
; GENERAL INFORMATION:
; APPLICANT: Koziel, Michael G.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Lewis, Kelly S.
; APPLICANT: Kramer, Vance C.
; APPLICANT: Warren, Gregory W.
; APPLICANT: Evola, Stephen V.
; APPLICANT: Crossland, Lyle D.
; APPLICANT: Wright, Martha S.
; APPLICANT: Merlin, Ellis J.
; APPLICANT: Launis, Karen L.
; APPLICANT: Rothstein, Steven J.
; APPLICANT: Bowman, Cindy G.
; APPLICANT: Dawson, John L.
; APPLICANT: Dunder, Erik M.
; APPLICANT: Pace, Gary M.
; APPLICANT: Suttie, Janet L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 601810artis Corporation
; STREET: Patent & Trademark Dept., 520 White Plains
; STREET: Rd., POB 2005
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-9005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,595A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40403
; REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8582
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3546 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3543
; OTHER INFORMATION: /product= "Full-length, hybrid,
; OTHER INFORMATION: maize optimized heat stable cryIA(b)"
; OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pCIB5513."
US-08-459-595A-14

Query Match   3.5%;   Score 35; DB 3; Length 3546;
Best Local Similarity   46.2%;   Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy    27 tcaacctcatccaccggacatcttcccctgctcaccagttccgctgtaagaaggagg 86
Db  2546 TCAAGATCAAGACCCAGGACGGCCCACGCGCCGCCTGGGCCAANTCTAGAGTTCCTGGAGCAGA 2605

Qy    87 gcccacccctcagtgtgccaatggttcaggtgaatgcctcctcaagtaccagctccgtc 146
Db  2606 ...
```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,448A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/951,715
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/772,027
; FILING DATE: 04-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40403
; REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8582
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3546 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3543
; OTHER INFORMATION: /product= "Full-length, hybrid,
; OTHER INFORMATION: maize optimized heat stable cryIA(b)"
; OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pCIB5513."
US-08-459-448A-14

Query Match   3.5%;   Score 35; DB 2; Length 3546;
Best Local Similarity   46.2%;   Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy    27 tcaacctcatccaccggacatcttcccctgctcaccagttccgctgtaagaaggagg 86
Db  2546 TCAAGATCAAGACCCAGGACGGCCCACGCGCCGCCTGGGCCAANTCTAGAGTTCCTGGAGCAGA 2605

Qy    87 gcccacccctcagtgtgccaatggttcaggtgaatgcctcctcaagtaccagctccgtc 146
Db  2606 AGCCCTGGTGGGCGGGCAGGCCCTGGCCCGCGGGCCGTGAACGCGGCGGAGAGAAGTGGCGCCACA 2665

Qy   147 ccaggaggagtgcgaggaggatgccattattacttgcaatcctgaggaattcatagttg 206
Db  2666 AGCCGGAGAAGCTGGAGTGGGAGACCAACAATCGTGTACAAGGAGGCCAAGGAGAGCGGTGG 2725

Qy   207 aggcgctgagctcccaacttccagcaggagtacagaggaggagttcgcagg 266
Db  2726 ACGCCCTGTTCGTGAACACCCCAGTACGACGCCGCTGCAGGCCGACACCACAACATCGCCATGA 2785

Qy   267 acggccagcc 277
Db  2786 TCCACGCCGCC 2796
```

```
Db  2606  AGCCCCTTGGTGGGCGGAGGCCCTTGGCCGCGTGAAGCCCGCCGAGAGAAAGTGGCGCGACA  2665
                                                                          
Qy   147  ccaggaggagtggcagaggatgccattattacttgcaatcctgaggaaattcaatagttg  206
                                                                          
Db  2666  AGCGCGAGAAGCTGGACTGGGAGACCAACATCGTGTACAAGAGGCCAAGGAGAGCGTGG  2725
                                                                          
Qy   207  agggcgctgcagcttcccaacttccagcagagagagtacagaggaggagtgcgcagg  266
                                                                          
Db  2726  ACGCCCTGTTCGTGAACAGCCAGTACGACCGCCTGCAGGCCGACACCAACATCGGCCATGA  2785
                                                                          
Qy   267  acggcccaagcc  277
                                                                          
Db  2786  TCCACGCGCGCC  2796
```

RESULT 6
US-08-459-504B-14
Sequence 14, Application US/0845950504B
Patent No. 6075185
GENERAL INFORMATION:
  APPLICANT: Koziel, Michael G.
  APPLICANT: Desai, Nalini M.
  APPLICANT: Lewis, Kelly S.
  APPLICANT: Kramer, Vance C.
  APPLICANT: Warren, Gregory W.
  APPLICANT: Evola, Stephen V.
  APPLICANT: Crossland, Lyle D.
  APPLICANT: Wright, Martha S.
  APPLICANT: Merlin, Ellis J.
  APPLICANT: Launis, Karen L.
  APPLICANT: Rothstein, Steven J.
  APPLICANT: Bowman, Cindy G.
  APPLICANT: Dawson, John L.
  APPLICANT: Dunder, Erik M.
  APPLICANT: Pace, Gary M.
  APPLICANT: Suttie, Janet L.
  TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
  TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
  NUMBER OF SEQUENCES: 94
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: NO. 6075185artis Corporation
    STREET: 3054 Cornwallis Road
    CITY: Research Triangle Park
    STATE: NC
    COUNTRY: USA
    ZIP: 27709
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/459,504B
    FILING DATE:
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/459,595
    FILING DATE: 02-JUN-1995
    APPLICATION NUMBER: US 07/951,715
    FILING DATE: 25-SEP-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/772,027
    FILING DATE: 04-OCT-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Meigs, J. Timothy
    REGISTRATION NUMBER: 38,241
    REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (919)541-8587
    TELEFAX: (919)541-8689
  INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 3546 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: other nucleic acid
  DESCRIPTION: /desc = "nucleic acid"
  HYPOTHETICAL: NO
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..3543
    OTHER INFORMATION: /product= "full-length, hybrid,
    OTHER INFORMATION: maize optimized heat stable cryIA(b)"
    OTHER INFORMATION: /note= "Disclosed in Figure 13 as contained in pCIB5513."
US-08-459-504B-14

Query Match   3.5%; Score 35; DB 3; Length 3546;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
Qy    27  tcaacctcatccacccggacatcttcccctgctcaccagtttccgctcgtaagaaggagg   86
                                                                            
Db  2546  TCAAGATCAAGACCCCAGGACACGGGCCCACCCCCGCTGGGCAATTCTAGAGTTCCTGAGAGA  2605
                                                                            
Qy    87  gcccaccctcagtgtgcccatggttcaggtgatgcctcctccaagtaccagctccgtc  146
                                                                            
Db  2606  AGCCCCTTGGTGGGCGGAGGCCCTTGGCCGCGTGAAGCCCGCCGAGAGAAAGTGGCGCGACA  2665
                                                                            
Qy   147  ccaggaggagtggcagaggatgccattattacttgcaatcctgaggaaattcaatagttg  206
                                                                            
Db  2666  AGCGCGAGAAGCTGGACTGGGAGACCAACATCGTGTACAAGAGGCCAAGGAGAGCGTGG  2725
                                                                            
Qy   207  agggcgctgcagcttcccaacttccagcagagagagtacagaggaggagtgcgcagg  266
                                                                            
Db  2726  ACGCCCTGTTCGTGAACAGCCAGTACGACCGCCTGCAGGCCGACACCAACATCGGCCATGA  2785
                                                                            
Qy   267  acggcccaagcc  277
                                                                            
Db  2786  TCCACGCGCGCC  2796
```

RESULT 7
US-08-459-444-14
Sequence 14, Application US/0845950444A
Patent No. 6121014
GENERAL INFORMATION:
  APPLICANT: Koziel, Michael G.
  APPLICANT: Desai, Nalini M.
  APPLICANT: Lewis, Kelly S.
  APPLICANT: Kramer, Vance C.
  APPLICANT: Warren, Gregory W.
  APPLICANT: Evola, Stephen V.
  APPLICANT: Crossland, Lyle D.
  APPLICANT: Wright, Martha S.
  APPLICANT: Merlin, Ellis J.
  APPLICANT: Launis, Karen L.
  TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
                      NUCLEIC ACID CODING SEQUENCE
  NUMBER OF SEQUENCES: 94
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.
    STREET: 3054 Cornwallis Road
    CITY: Research Triangle Park
    STATE: NC
    COUNTRY: USA
    ZIP: 27709
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/459,444A
    FILING DATE: 02-Jun-1995

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/951,715
  FILING DATE: 25-SEP-1992
  APPLICATION NUMBER: US 07/772,027
  FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
  NAME: Meigs, J. Timothy
  REGISTRATION NUMBER: 38,241
  REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (919)541-8689
  TELEFAX: (919)541-8587
INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 3546 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: other nucleic acid
    DESCRIPTION: /desc = "Synthetic DNA"
  HYPOTHETICAL: NO
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..3543
    OTHER INFORMATION: /product= "Full-length, hybrid,
    maize optimized heat stable cryIA(b)"
    /note= "Disclosed in Figure 13 as contained in pCIB5513."
  SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-08-459-444-14

Query Match   3.5%; Score 35; DB 3; Length 3546;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy   27 tcaacctcatccaccccggacatcttccccctgctccaccagtttccgctgtaagaaggagg   86
Db 2546 TCAAGATCAAGACCCAGGACGGACCCACGCCCCCCTGGGCAATCTAGAGTTCCTTGGAGGAGA 2605

Qy   87 gcccacccctcagtgtgtgccccatggttcaggtgaatgcctcctcaagtaccagctccgtc 146
Db 2606 AGCCCCTGGTGGGCGGAGGCCCTGCCCCGTGAAGCCGCCGAGAGAAGAGTTGGCGGCGACA 2665

Qy   147 ccaggaggagtggcagaggatgccattattacttgcaatcctgagggaattcatagttg 206
Db 2666 AGCGGGAGAAGCTGGAGTGGGGAGACCAAACATCGTGTACAAGGAGGGCCAAGGAGGGTGG 2725

Qy   207 aggcggtgcaagcttcccaacttccagcagagcagtcgtcaggagtacaggagggagtgcggcagg 266
Db 2726 ACGCCCTGTTCGTGAACAGCCAGTACGACGCCGACACCACAACATCGCCATGA 2785

Qy   267 acggcccagcc 277
Db 2786 TCCACGCGCGCC 2796

RESULT   8
US-09-053-549-1
Sequence 1, Application US/0905349
Patent No. 6121521
GENERAL INFORMATION:
APPLICANT: Desai, Nalini
TITLE OF INVENTION: Insecticidal Protein and Gene
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
  ADDRESSEE: No. 612152Iartis Corporation
  STREET: 3054 Cornwallis Rd.
  CITY: Research Triangle Park
  STATE: NC
  COUNTRY: USA
  ZIP: 27709
  COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/053,549
  FILING DATE: 01-APR-1998
  CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
  NAME: Pace, Gary M.
  REGISTRATION NUMBER: 40,403
  REFERENCE/DOCKET NUMBER: CGC 1995
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 919-541-8689
  TELEFAX: 919-541-8582
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 8854 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  HYPOTHETICAL: NO
  FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 8854
    OTHER INFORMATION: /product= "hyFLIB protein"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 11..3694
    OTHER INFORMATION: /note= "Sequence of pCIB5520
    containing coding sequence for hyFLIB protein"
Patent No. 6121521
US-09-053-549-1

Query Match   3.5%; Score 35; DB 3; Length 8854;
Best Local Similarity 46.2%; Pred. No. 1.7;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy   27 tcaacctcatccaccccggacatcttccccctgctccaccagtttccgctgtaagaaggagg   86
Db 2634 TCAAGATCAAGACCCAGGACGGACCCACGCCCCCCTGGGCAATCTAGAGTTCCTTGGAGGAGA 2693

Qy   87 gcccacccctcagtgtgtgccccatggttcaggtgaatgcctcctcaagtaccagctccgtc 146
Db 2694 AGCCCCTGGTGGGCGGAGGCCCTGCCCCGTGAAGCCGCCGAGAGAAGAGTTGGCGGCGACA 2753

Qy   147 ccaggaggagtggcagaggatgccattattacttgcaatcctgagggaattcatagttg 206
Db 2754 AGCGGGAGAAGCTGGAGTGGGGAGACCAAACATCGTGTACAAGGAGGGCCAAGGAGGGTGG 2813

Qy   207 aggcggtgcaagcttcccaacttccagcagagcagtcgtcaggagtacaggagggagtgcggcagg 266
Db 2814 ACGCCCTGTTCGTGAACAGCCAGTACGACGCCGACACCACAACATCGCCATGA 2873

Qy   267 acggcccagcc 277
Db 2874 TCCACGCGCGCC 2933

RESULT   9
US-07-951-715A-2
Sequence 2, Application US/07951715A
Patent No. 5625136
GENERAL INFORMATION:
APPLICANT: Koziel, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.

```
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES: 94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: CIBA-GEIGY Corporation
;      STREET: 7 Skyline Drive
;      CITY: Hawthorne
;      STATE: New York
;      COUNTRY: USA
;      ZIP: 10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: Patentin Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/07/951,715A
;      FILING DATE: 25-SEP-1992
;      CLASSIFICATION: 800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/772,027
;      FILING DATE: 04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Spruill, W. Murray
;      REGISTRATION NUMBER: 32,943
;      REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (919)541-8689
;      TELEFAX: (919)541-8615
;    INFORMATION FOR SEQ ID NO: 2:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: 3468 base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;      MOLECULE TYPE: other nucleic acid
;      DESCRIPTION: /desc = "Synthetic DNA"
;      HYPOTHETICAL: NO
;      FEATURE:
;        NAME/KEY: misc_feature
;        LOCATION: 1..3468
;        OTHER INFORMATION: /product= "Full-length pure maize
;        OTHER INFORMATION: optimized synthetic Bt"
;        OTHER INFORMATION: /note= "Disclosed in Figure 3 as synlt.mze"
     US-07-951-715A-2
```

```
Query Match      Best Local Similarity  45.8%;          Score 33.4;  DB 1;  Length 3468;
Matches 115;  Conservative 0;  Mismatches 136;  Indels 0;  Gaps 0;
```

```
Qy    27 tcaacctcatccaccccggacatcttccccctgctcaccagttccgctgtaaagaaggagg   86
         ||||  |||| ||    |||||||  ||   |||  |  |||||  | ||||  |||||||
Db  2468 TCAAGATCAAGAGACCCAGGACGCCACGCGCCCACGCGCCTGGCGCACACCTGGAGTTCCTGGAGGAGA  2527

Qy    87 gcccacccctcagttgtgccccatcggttcaaggtgaatggcctcctcaagtaccagctccgtc  146
         ||||  |||||  |  |||||||||  ||||||||||  |||  ||||  ||||| |||||
Db  2528 AGCCCCTGGTGGCGCGAGGCCCTGGCCCGCCGTGAAAGCCGCCGAGAACGAAGTGCGCGCGACA  2587

Qy   147 ccaggaggagtggcagaggaggatgccattattacttgcaatcctgaggaattcaatagttg  206
         |||  |||||||  |||||||||| |||  ||||||  | |||| ||||||
Db  2588 AGCGCGAGAAGCGTGGACTGGGAGACCAACATCGTGTACAAGGAGGCCAAGGAGAGCGGTGG  2647

Qy   207 aggcgctgcagcttccaacttccagcagagcgtgcaggagtacaggaggagtgcgcagg  266
         |||||||||  |||||||||||  ||||||||  ||||||||  |||||||||||||||
```

```
RESULT 10
US-07-951-715A-4
Sequence 4, Application US/07951715A
Patent No. 5625136
GENERAL INFORMATION:
;    APPLICANT:  Koziel, Michael G.
;    APPLICANT:  Desai, Nalini M.
;    APPLICANT:  Lewis, Kelly S.
;    APPLICANT:  Kramer, Vance C.
;    APPLICANT:  Warren, Gregory W.
;    APPLICANT:  Evola, Stephen V.
;    APPLICANT:  Crossland, Lyle D.
;    APPLICANT:  Wright, Martha S.
;    APPLICANT:  Merlin, Ellis J.
;    APPLICANT:  Launis, Karen L.
;    APPLICANT:  Rothstein, Steven J.
;    APPLICANT:  Bowman, Cindy G.
;    APPLICANT:  Dawson, John L.
;    APPLICANT:  Dunder, Erik M.
;    APPLICANT:  Pace, Gary M.
;    APPLICANT:  Suttie, Janet L.
;    TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
;    TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
;    NUMBER OF SEQUENCES: 94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: CIBA-GEIGY Corporation
;      STREET: 7 Skyline Drive
;      CITY: Hawthorne
;      STATE: New York
;      COUNTRY: USA
;      ZIP: 10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: Patentin Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/07/951,715A
;      FILING DATE: 25-SEP-1992
;      CLASSIFICATION: 800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 07/772,027
;      FILING DATE: 04-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Spruill, W. Murray
;      REGISTRATION NUMBER: 32,943
;      REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (919)541-8689
;      TELEFAX: (919)541-8615
;    INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: 3468 base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;      MOLECULE TYPE: other nucleic acid
;      DESCRIPTION: /desc = "Synthetic DNA"
;      HYPOTHETICAL: NO
;      FEATURE:
;        NAME/KEY: misc_feature
;        LOCATION: 1..3468
;        OTHER INFORMATION: /product= "Full length synthetic
;        OTHER INFORMATION: maize optimized"
;        OTHER INFORMATION: /note= "Disclosed in Figure 3 as synful.mod. This sequenc
```

```
Db  2648 ACGCCCTGTTCGTGAACAGCCAGTACGACCGGCCTGCAGCGCCGACCACCAACATCGCCATGA  2707
         ||| |||| ||  |||  | | || ||||  ||  |  | ||| ||  |||| | ||| | ||
Qy   267 acggcccagc  277
         ||  |||||
Db  2708 TCCACGCCGCC  2718
```

OTHER INFORMATION: identical to flsynbt.fin as disclosed in Figure 1."

RESULT 11
US-08-459-448A-2
Sequence 2, Application US/08459448A
Patent No. 5859336
GENERAL INFORMATION:

```
Query Match
Best Local Similarity   45.8%;    Score 33.4;  DB 1;  Length 3468;
Matches 115;  Conservative   0;  Mismatches 136;  Indels   0;  Gaps   0;

Qy    27 tcaacctcatccaccggacatcttcccctgctcaccagttccgctgtaagaaggagg   86
            ||||  ||| ||   ||| ||     ||| || || ||   |||    |||  |||||
Db  2468 TCAAGATCAAGAACCCAGGACGACGGCCACGCGCCGCTGGGCAACCTGGAGTTCCTGGAGGAGA   2527

Qy    87 gccccacctcagtgtgccatggttcagggtgaatgcctcctcaagtaccagctccgtc   146
            ||  |  |||  || ||| |||| ||| || || ||||   |||  ||   ||  ||
Db  2528 AGCCCTGGTGGGCGAGGCCCTGGCCCCGCGCCGCCGTGAACGCGCCCGAGAAGAAGTGCGCGACA   2587

Qy   147 ccaggaggagtggcaagaggagatggcattattacttgcaatcctgaggaattcatagttg   206
            ||  |||||| ||| || ||  ||   ||| ||| ||  ||  || ||  |||  || |||
Db  2588 AGCCGGAGAAGCTGGAGTGGGAGACCAACATCGTGTACAAGGAGGCCAAGGACGCGGTGG   2647

Qy   207 aggcgctgcagcttcccaacttccagcagagcgtgcaggagtacaggagggagtcgcagg   266
            ||   |||||| |||    |||  || || || |||     ||  ||  ||  ||  |||
Db  2648 ACGGCCTGTTCGTGAACAGCCAGTACGACCGCCTCAGGCCGCGAACACCAACATCGCCATGA   2707

Qy   267 acggcccagcc   277
            ||  |  ||||
Db  2708 TCCACGCGCGC   2718
```

RESULT 12
US-08-459-448A-4
Sequence 4, Application US/08459448A
Patent No. 5859336
GENERAL INFORMATION:

```
Query Match
Best Local Similarity   45.8%;    Score 33.4;  DB 2;  Length 3468;
Matches 115;  Conservative   0;  Mismatches 136;  Indels   0;  Gaps   0;

Qy    27 tcaacctcatccaccggacatcttcccctgctcaccagttccgctgtaagaaggagg   86
            ||||  ||| ||   ||| ||     ||| || || ||   |||    |||  |||||
Db  2468 TCAAGATCAAGAACCCAGGACGACGGCCACGCGCCGCTGGGCAACCTGGAGTTCCTGGAGGAGA   2527

Qy    87 gccccacctcagtgtgccatggttcagggtgaatgcctcctcaagtaccagctccgtc   146
            ||  |  |||  || ||| |||| ||| || || ||||   |||  ||   ||  ||
Db  2528 AGCCCTGGTGGGCGAGGCCCTGGCCCCGCGCCGCCGTGAACGCGCCCGAGAAGAAGTGCGCGACA   2587

Qy   147 ccaggaggagtggcaagaggagatggcattattacttgcaatcctgaggaattcatagttg   206
            ||  |||||| ||| || ||  ||   ||| ||| ||  ||  || ||  |||  || |||
Db  2588 AGCCGGAGAAGCTGGAGTGGGAGACCAACATCGTGTACAAGGAGGCCAAGGACGCGGTGG   2647

Qy   207 aggcgctgcagcttcccaacttccagcagagcgtgcaggagtacaggagggagtcgcagg   266
            ||   |||||| |||    |||  || || || |||     ||  ||  ||  ||  |||
Db  2648 ACGGCCTGTTCGTGAACAGCCAGTACGACCGCCTCAGGCCGCGAACACCAACATCGCCATGA   2707

Qy   267 acggcccagcc   277
            ||  |  ||||
Db  2708 TCCACGCGCGC   2718
```

PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/772,027
    FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
    NAME: Pace, Gary M.
    REGISTRATION NUMBER: 40403
    REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (919)541-8582
    TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
    LENGTH: 3468 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
    DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 1..3468
    OTHER INFORMATION: /product= "Full-length pure maize
    OTHER INFORMATION: optimized synthetic Bt"
    OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn1T.mze"
US-08-459-448A-2

Best Local Similarity   45.8%;
Matches 115; Conservative   0;
US-08-459-448A-2

CORRESPONDENCE ADDRESS:
    ADDRESSEE: No. 5859336artis Corporation
    STREET: Patent & Trademark Dept., 520 White Plains
    STREET: Rd., POB 2005
    CITY: Tarrytown
    STATE: New York
    COUNTRY: USA
    ZIP: 10591-9005
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/459,448A
    FILING DATE: 02-JUN-1995
    CLASSIFICATION: 800
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/951,715
    FILING DATE: 25-SEP-1992

TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94

APPLICANT: Koziel, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttie, Janet L.

```
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
   ADDRESSEE: No. 5859336artis Corporation
   STREET: Patent & Trademark Dept., 520 White Plains
   STREET: Rd., POB 2005
   CITY: Tarrytown
   STATE: New York
   COUNTRY: USA
   ZIP: 10591-9005
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US 07/772,027
   FILING DATE: 04-OCT-1991
   APPLICATION NUMBER: US/08/459,448A
   FILING DATE: 02-JUN-1995
   CLASSIFICATION: 800
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 07/951,715
   FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 07/772,027
   FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
   NAME: Pace, Gary M.
   REGISTRATION NUMBER: 40403
   REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (919)541-8582
   TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 4:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 3468 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: other nucleic acid
   DESCRIPTION: /desc = "Synthetic DNA"
   HYPOTHETICAL: NO
   FEATURE:
      NAME/KEY: misc_feature
      LOCATION: 1..3468
      OTHER INFORMATION: /product= "Full length synthetic
      OTHER INFORMATION: maize optimized"
      OTHER INFORMATION: /note= "Disclosed in Figure 3 as synful.mod. This sequence i
      DESCRIPTION: identical to flsynbt.fin as disclosed in Figure 1."
US-08-459-448A-4
```

```
Query Match        3.3%;  Score 33.4;  DB 2;  Length 3468;
Best Local Similarity 45.8%;   Pred. No. 3.1;
Matches 115;  Conservative 0;  Mismatches 136;  Indels 0;  Gaps 0;

Qy   27 tcaacttcatccaaccggacatcttccccctgctcaacagttccgctgtaagagaagg    86
         ||| ||||| |||| |  ||||||||||| ||||||| | ||||||
Db 2468 TCAAGATCAAGACCAGGAGCGGCACGGCCACCGCCTGGGCAACCTGGAGTTCCTGGAGGAGA 2527

Qy   87 gcccacccctcagtgtgcccatggttcaggggtgaatgcctcctcaagtaccagctccgtc  146
          ||| ||  || ||||||  || ||
Db 2528 AGCCCTGTGGGCGAGGCCCTGGCCCGCCCGTGAAGCCCGCCAGAAAGAAGTGCGCGACA   2587

Qy  147 ccaggaggagtggcaggagggatggccattattacttgcaatcctgaggaattcatagttg  206
Db 2588 AGCGGCGAGAAGCTGGAGTGGGAGACCAACATCGTGTACAAGGAGGCCAAGGAGAGCCGTGG 2647

Qy  207 aggcgctgcaggtcccaacttccagcagaggagtacacaggaggagtgcgcagg        266
Db 2648 ACGCCCGTGTTCGTGAACAGCCAGTACGACCCGCCTGCAGCCGACACCAACATCGCCATGA 2707

Qy  267 acggcccagcc  277
         |||||| |||
Db 2708 TCCACCGCCGCC 2718
```

```
RESULT 13
US-08-459-595A-2
Sequence 2, Application US/08459595A
Patent No. 6018104
GENERAL INFORMATION:
   APPLICANT: Koziel, Michael G.
   APPLICANT: Desai, Nalini M.
   APPLICANT: Lewis, Kelly S.
   APPLICANT: Kramer, Vance C.
   APPLICANT: Warren, Gregory W.
   APPLICANT: Evola, Stephen V.
   APPLICANT: Crossland, Lyle D.
   APPLICANT: Wright, Martha S.
   APPLICANT: Merlin, Ellis J.
   APPLICANT: Launis, Karen L.
   APPLICANT: Rothstein, Steven J.
   APPLICANT: Bowman, Cindy G.
   APPLICANT: Dawson, John L.
   APPLICANT: Dunder, Erik M.
   APPLICANT: Pace, Gary M.
   APPLICANT: Suttie, Janet L.
   TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
   TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
   NUMBER OF SEQUENCES: 94
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: No. 6018104artis Corporation
      STREET: Patent & Trademark Dept., 520 White Plains
      STREET: Rd., POB 2005
      CITY: Tarrytown
      STATE: New York
      COUNTRY: USA
      ZIP: 10591-9005
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/459,595A
      FILING DATE: 02-JUN-1995
      CLASSIFICATION: 800
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/951,715
      FILING DATE: 25-SEP-1992
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/772,027
      FILING DATE: 04-OCT-1991
   ATTORNEY/AGENT INFORMATION:
      NAME: Pace, Gary M.
      REGISTRATION NUMBER: 40403
      REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: (919)541-8582
      TELEFAX: (919)541-8689
   INFORMATION FOR SEQ ID NO: 2:
      SEQUENCE CHARACTERISTICS:
         LENGTH: 3468 base pairs
         TYPE: nucleic acid
         STRANDEDNESS: single
         TOPOLOGY: linear
      MOLECULE TYPE: other nucleic acid
      DESCRIPTION: /desc = "Synthetic DNA"
      HYPOTHETICAL: NO
      FEATURE:
         NAME/KEY: misc_feature
         LOCATION: 1..3468
         OTHER INFORMATION: /product= "Full-length pure maize
         OTHER INFORMATION: optimized synthetic Bt."
```

```
; OTHER INFORMATION: /note= "Disclosed in Figure 3 as synIT.mze"
US-08-459-595A-2

Query Match      Best Local Similarity   45.8%;   Score 33.4;  DB 3;  Length 3468;
Matches 115;  Conservative   0;  Mismatches  136;  Indels   0;  Gaps   0;

Qy      27  tcaacctcatccacccggacatcttccccctgctcaccagttccgctgtaagagaggg  86
            |||| ||| ||| ||||||| ||| ||||||||||||| ||| ||| ||  |||||||
Db    2468  TCAAGATCAAGACCCAGGACACGGCGCCCGCCCTGGGCAACCTGGAGTTCCTGGAGGAGA  2527

Qy      87  gccccaccctcagtgtgcccatggttcaggtgaatgcctcctcaagtaccagctccgtc  146
            |||| |||  |||||||||||||||| |||||  |||||  ||||||||||| ||| |||
Db    2528  AGCCCTGGTGTGGGCGGAGGCCCTGGCCCGGTGAAGCGCGCCGGCGAAGAAGTGGCGGACA  2587

Qy     147  ccaggaggagtggcaggaggatgccattattacttgcaatcctgaggaattcatagttg  206
            |||| |||  ||| ||| || ||||| ||  ||||||||  ||| ||||| ||||| |||
Db    2588  AGCCGGAGAAGCTGGAGTGGGAGACCAACATCGTGTACAAGGAGAGCCAAGGAGAGAGCGTGG  2647

Qy     207  aggcgctgcagcttcccaacttccagcacgagacgtgcaggaggtacaggagaggtgcgcagg  266
            |||| ||| |||  |||  ||||  ||| ||  |||||| |||  ||| ||||  |  |||
Db    2648  ACGCCCTGTTCGTGAACAGCCAGTACGACCCGTCGCAGCCGCAGCACCAACATCGCCATGA  2707

Qy     267  acggcccagcc  277
            ||  |||| |||
Db    2708  TCCACGCCGCC  2718
```

```
RESULT 14
US-08-459-595A-4
; Sequence 4, Application US/0845595A
; Patent No. 6018104
GENERAL INFORMATION:
    APPLICANT: Koziel, Michael G.
    APPLICANT: Desai, Nalini M.
    APPLICANT: Lewis, Kelly S.
    APPLICANT: Kramer, Vance C.
    APPLICANT: Warren, Gregory W.
    APPLICANT: Evola, Stephen V.
    APPLICANT: Crossland, Lyle D.
    APPLICANT: Wright, Martha S.
    APPLICANT: Merlin, Ellis J.
    APPLICANT: Launis, Karen L.
    APPLICANT: Rothstein, Steven J.
    APPLICANT: Bowman, Cindy G.
    APPLICANT: Dawson, John L.
    APPLICANT: Dunder, Erik M.
    APPLICANT: Pace, Gary M.
    APPLICANT: Suttie, Janet L.
    TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
    TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
    NUMBER OF SEQUENCES: 94
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: No. 6018104artis Corporation
        STREET: Patent & Trademark Dept., 520 White Plains
        STREET: Rd., POB 2005
        CITY: Tarrytown
        STATE: New York
        COUNTRY: USA
        ZIP: 10591-9005
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/459,595A
        FILING DATE: 02-JUN-1995
        CLASSIFICATION: 800
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/951,715
        FILING DATE: 25-SEP-1992
```

```
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/772,027
    FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
    NAME: Pace, Gary M.
    REGISTRATION NUMBER: 40403
    REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (919)541-8582
    TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 4:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 3468 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
    MOLECULE TYPE: other nucleic acid
    DESCRIPTION: /desc = "Synthetic DNA"
    HYPOTHETICAL: NO
    FEATURE:
        NAME/KEY: misc_feature
        LOCATION: 1..3468
        OTHER INFORMATION: /product= "Full length synthetic
        OTHER INFORMATION: maize optimized"
        OTHER INFORMATION: /note= "Disclosed in Figure 3 as synful.mod. This sequenc
        OTHER INFORMATION: identical to flsynbt.fin as disclosed in Figure 1."
US-08-459-595A-4
```

```
Query Match      Best Local Similarity   45.8%;   Score 33.4;  DB 3;  Length 3468;
Matches 115;  Conservative   0;  Mismatches  136;  Indels   0;  Gaps   0;

Qy      27  tcaacctcatccacccggacatcttccccctgctcaccagttccgctgtaagagaggg  86
            |||| ||| ||| ||||||| ||| ||||||||||||| ||| ||| ||  |||||||
Db    2468  TCAAGATCAAGACCCAGGACACGGCGCCCGCCCTGGGCAACCTGGAGTTCCTGGAGGAGA  2527

Qy      87  gccccaccctcagtgtgcccatggttcaggtgaatgcctcctcaagtaccagctccgtc  146
            |||| |||  |||||||||||||||| |||||  |||||  ||||||||||| ||| |||
Db    2528  AGCCCTGGTGTGGGCGGAGGCCCTGGCCCGGTGAAGCGCGCCGGCGAAGAAGTGGCGGACA  2587

Qy     147  ccaggaggagtggcaggaggatgccattattacttgcaatcctgaggaattcatagttg  206
            |||| |||  ||| ||| || ||||| ||  ||||||||  ||| ||||| ||||| |||
Db    2588  AGCCGGAGAAGCTGGAGTGGGAGACCAACATCGTGTACAAGGAGAGCCAAGGAGAGAGCGTGG  2647

Qy     207  aggcgctgcagcttcccaacttccagcacgagacgtgcaggaggtacaggagaggtgcgcagg  266
            |||| ||| |||  |||  ||||  ||| ||  |||||| |||  ||| ||||  |  |||
Db    2648  ACGCCCTGTTCGTGAACAGCCAGTACGACCCGTCGCAGCCGCAGCACCAACATCGCCATGA  2707

Qy     267  acggcccagcc  277
            ||  |||| |||
Db    2708  TCCACGCCGCC  2718
```

```
RESULT 15
US-08-459-504B-2
; Sequence 2, Application US/0845595504B
; Patent No. 6075185
GENERAL INFORMATION:
    APPLICANT: Koziel, Michael G.
    APPLICANT: Desai, Nalini M.
    APPLICANT: Lewis, Kelly S.
    APPLICANT: Kramer, Vance C.
    APPLICANT: Warren, Gregory W.
    APPLICANT: Evola, Stephen V.
    APPLICANT: Crossland, Lyle D.
    APPLICANT: Wright, Martha S.
    APPLICANT: Merlin, Ellis J.
    APPLICANT: Launis, Karen L.
    APPLICANT: Rothstein, Steven J.
    APPLICANT: Bowman, Cindy G.
    APPLICANT: Dawson, John L.
    APPLICANT: Dunder, Erik M.
```

~ APPLICANT: Pace, Gary M.
~ APPLICANT: Suttie, Janet L.
~ TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
~ TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
~ NUMBER OF SEQUENCES: 94
~ CORRESPONDENCE ADDRESS:
~   ADDRESSEE: NO. 6075185artis Corporation
~   STREET: 3054 Cornwallis Road
~   CITY: Research Triangle Park
~   STATE: NC
~   COUNTRY: USA
~   ZIP: 27709
~ COMPUTER READABLE FORM:
~   MEDIUM TYPE: Floppy disk
~   COMPUTER: IBM PC compatible
~   OPERATING SYSTEM: PC-DOS/MS-DOS
~   SOFTWARE: PatentIn Release #1.0, Version #1.30
~ CURRENT APPLICATION DATA:
~   APPLICATION NUMBER: US/08/459,504B
~   FILING DATE:
~ PRIOR APPLICATION DATA:
~   CLASSIFICATION:
~   APPLICATION NUMBER: US 08/459,595
~   FILING DATE: 02-JUN-1995
~   APPLICATION NUMBER: US 07/951,715
~   FILING DATE: 25-SEP-1992
~ PRIOR APPLICATION DATA:
~   APPLICATION NUMBER: US 07/772,027
~   FILING DATE: 04-OCT-1991
~ ATTORNEY/AGENT INFORMATION:
~   NAME: Meigs, J. Timothy
~   REGISTRATION NUMBER: 38, 241
~   REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV
~ TELECOMMUNICATION INFORMATION:
~   TELEPHONE: (919)541-8587
~   TELEFAX: (919)541-8689
~ INFORMATION FOR SEQ ID NO: 2:
~ SEQUENCE CHARACTERISTICS:
~   LENGTH: 3468 base pairs
~   TYPE: nucleic acid
~   STRANDEDNESS: single
~   TOPOLOGY: linear
~ MOLECULE TYPE: other nucleic acid
~ DESCRIPTION: /desc = "Synthetic DNA"
~ HYPOTHETICAL: NO
~ FEATURE:
~   NAME/KEY: misc_feature
~   LOCATION: 1..3468
~   OTHER INFORMATION: /product= "Full-length pure maize
~   OTHER INFORMATION: optimized synthetic Bt"
~   OTHER INFORMATION: /note= "Disclosed in Figure 3 as syn1T.mze"
US-08-459-504B-2

Query Match   3.3%;   Score 33.4; DB 3; Length 3468;
Best Local Similarity 45.8%;   Pred. No. 3.1;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy   27 tcaacctcatccaccggaacatcttcccctgctccaccagttccgctgtaagaaggagg 86
         |||| |||| |||| |||| ||| |||| | |||| || ||| || ||||||
Db 2468 TCAAGATCAAGACCACCAGGACGGCCACCGCCTGGGCAACCTGGAGTTCCTGGAGGAGA 2527

Qy   87 gcccacccctcaagtgtgcccatggttcaggtgaatgcctcctcaagtaccagctccgtc 146
         |.| |||| | ||| |||| |||| || |||| ||||| ||| |
Db 2528 AGCCCTGGTGGGCGAGGCCCTGGCCCCGCGCGTGAAGCGCCGAGAAGAAGTGGCGCGACA 2587

Qy   147 ccaggagggagtggcagaggatgcccattattacttgcaatcctgaggaattcaatagttg 206
         || ||||| | || || |||| ||| ||| ||| | | ||| ||||| ||| ||
Db 2588 AGCGCGAGAAGCTGGAGTGGGGAGACCAACATCGTGTACAAGGAGGGCCAAAGGAGGCGTGG 2647

Qy   207 agggcgctgcagcttcccaacttccagcagaggaggcgtgcaggaggagtgcgcgagg 266
         ||| ||| | || ||| | ||| |||| | |||||| |||| | |||
Db 2648 ACGCCCTGCTTCGTGAACAGCCAGTACCAGCGCTCGCAGCCCGACACACAACATCGCCATGA 2707

Qy   267 acggcccagcc 277
         | | | |||
Db 2708 TCCACGCCGCC 2718

Search completed: October 28, 2001, 21:48:02
Job time: 2650 sec

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          October 28, 2001, 19:42:27 ; Search time 1157.4 Seconds
                                              (without alignments)
                                              8175.485 Million cell updates/sec

Title:           SEQLINS_COPY_1140_2140
Perfect score:   1001
Sequence:        1 tcgcagccacaagattcaaa..........aagccatcagcgtgggatg 1001

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters:    20456230

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :  EST:*

  1:  gb_est1:*               44:  em_esthum10:*
  2:  gb_est2:*               45:  em_esthum11:*
  3:  gb_est3:*               46:  em_esthum12:*
  4:  gb_est4:*               47:  em_esthum13:*
  5:  gb_est5:*               48:  em_esthum14:*
  6:  gb_est6:*               49:  em_esthum15:*
  7:  gb_est7:*               50:  em_esthum16:*
  8:  gb_est8:*               51:  em_esthum17:*
  9:  gb_est9:*               52:  em_esthum18:*
 10:  gb_est10:*              53:  em_esthum19:*
 11:  gb_est11:*              54:  em_esthum20:*
 12:  gb_est12:*              55:  em_esthum21:*
 13:  gb_est13:*              56:  em_esthum22:*
 14:  gb_est14:*              57:  em_esthum23:*
 15:  gb_est15:*              58:  em_esthum24:*
 16:  gb_est16:*              59:  em_esthum25:*
 17:  gb_est17:*              60:  em_esthum26:*
 18:  gb_est18:*              61:  em_esthum27:*
 19:  gb_est19:*              62:  em_esthum28:*
 20:  gb_est20:*              63:  em_estin1:*
 21:  gb_est21:*              64:  em_estin2:*
 22:  gb_est22:*              65:  em_estin3:*
 23:  gb_est23:*              66:  em_estin4:*
 24:  gb_est24:*              67:  em_estin5:*
 25:  gb_est33:*              68:  em_estom1:*
 26:  gb_est34:*              69:  em_estom2:*
 27:  gb_est35:*              70:  em_estov1:*
 28:  gb_est36:*              71:  em_estov2:*
 29:  gb_est37:*              72:  em_estpl1:*
 30:  gb_est38:*              73:  em_estpl2:*
 31:  gb_est39:*              74:  em_estpl3:*
 32:  gb_est40:*              75:  em_estpl4:*
 33:  em_estba:*              76:  em_estpl5:*
 34:  em_estfun:*             77:  em_estpl6:*
 35:  em_esthum1:*            78:  em_estpl7:*
 36:  em_esthum2:*            79:  em_estpl8:*
 37:  em_esthum3:*            80:  em_estpl9:*
 38:  em_esthum4:*            81:  em_estpl10:*
 39:  em_esthum5:*            82:  em_estro1:*
 40:  em_esthum6:*            83:  em_estro2:*
 41:  em_esthum7:*            84:  em_estro3:*
 42:  em_esthum8:*            85:  em_estro4:*
 43:  em_esthum9:*            86:  em_estro5:*
                             87:  em_estro6:*
                             88:  em_estro7:*
                             89:  em_estro8:*
                             90:  em_estro9:*
                             91:  em_estro10:*
                             92:  em_estro11:*
                             93:  em_estro12:*
                             94:  em_estro13:*
                             95:  em_estro14:*
                             96:  em_estro15:*
                             97:  em_estro16:*
                             98:  em_estro17:*
                             99:  em_estro18:*
                            100:  em_estro19:*
                            101:  em_estro20:*
                            102:  gb_est25:*
                            103:  gb_est26:*
                            104:  gb_est27:*
                            105:  gb_est28:*
                            106:  gb_est29:*
                            107:  gb_est30:*
                            108:  gb_est31:*
                            109:  gb_est32:*
                            110:  gb_est41:*
                            111:  gb_est42:*
                            112:  gb_est43:*
                            113:  gb_est44:*
                            114:  gb_est45:*
                            115:  gb_est46:*
                            116:  gb_est47:*
```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estp111:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*

190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun1:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_Other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

RESULT 1

LOCUS   BF797306   960 bp   mRNA   EST   12-JAN-2001
DEFINITION 602256926F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4340286 5',
            mRNA sequence.
ACCESSION   BF797306
VERSION   BF797306.1 GI:12102360
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 960)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.

FEATURES   Location/Qualifiers
     source   1..960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4340286"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT   230 a   286 c   276 g   168 t
ORIGIN

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	695	69.4	960	169	BF797306	BF797306 602256926
2	675.8	67.5	1510	192	AK004136	AK004136 Mus muscu
3	644.6	64.4	735	141	BE902696	BE902696 601677393
4	636.8	63.6	947	150	BF525432	BF525432 602005917
5	636.6	63.6	761	140	BF795820	BF795820 601590856
6	618	61.7	676	165	BE260495	BE260495 601150702
7	599.8	59.9	692	164	BE382353	BE382353 601298656
8	576.4	57.6	884	154	BG519751	BG519751 602578902
9	568.4	56.8	745	141	BE900936	BE900936 601674206
10	567.6	56.7	614	152	BG335963	BG335963 602404708
11	564.8	56.4	664	156	BE383336	BE383336 601298249
12	562.6	56.2	713	136	BE537086	BE537086 601063325
13	541.4	54.1	823	138	BE619259	BE619259 601473130
14	535.4	53.5	915	147	BF311926	BF311926 601897767
15	505.6	50.5	992	139	BE747163	BE747163 601577254
16	490.2	49.0	574	165	BE250309	BE250309 600434355
17	488.6	48.8	1497	147	BF315158	BF315158 601899518
18	470	47.0	878	147	BF338283	BF338283 602035507
19	454.8	45.4	920	153	BG386090	BG386090 602455264
20	450	45.0	1012	153	BG386348	BG386348 602455550
21	429.4	42.9	628	145	BF181255	BF181255 601805739
c 22	424.2	42.4	512	112	AW175923	AW175923 RC1-BT007
23	410.2	41.0	762	107	AU124671	AU124671 AU124671
c 24	389.2	38.9	852	146	BF240253	BF240253 601905776
25	383.8	38.3	961	153	BG387104	BG387104 601435738
26	380.2	38.0	790	140	BE892893	BE892893 601455813
27	379.4	37.9	698	140	BE795434	BE795434 601592991
28	377.2	37.7	617	141	BE914535	BE914535 601665633
29	369	36.9	543	31	AV615441	AV615441 AV615441
30	364	36.4	482	187	R55841	R55841 yg89d01.r1
31	362.8	36.2	1124	147	BF342802	BF342802 602015077
32	358.2	35.8	803	166	BE371842	BE371842 601217643
33	354	35.4	501	5	AA311855	AA311855 EST182568
34	352.2	35.2	472	187	R51138	R51138 yg71c08.r1
c 35	341.8	34.1	612	166	BE304720	BE304720 601106236
c 36	338.6	33.8	346	122	AW889463	AW889463 RC6-NT002
37	338	33.8	394	6	AA346268	AA346268 EST52407
38	335.4	33.5	589	108	AU139902	AU139902 AU139902
39	329.2	32.9	891	175	BG258460	BG258460 602379938
40	324.8	32.4	547	144	BF076468	BF076468 226041 MA
41	321	32.1	677	139	BE742908	BE742908 601574609
c 42	313.2	31.3	517	104	AI991599	AI991599 ws18c04.x
43	308.8	30.8	443	138	BE654091	BE654091 UI-M-AH1-
44	305.6	30.5	1112	192	AK003759	AK003759 Mus muscu
45	297	29.7	341	188	T16931	T16931 NIB2016-5R

ALIGNMENTS

Query Match   69.4%;   Score 695;   DB 169;   Length 960;
Best Local Similarity   96.8%;
Matches 752; Conservative   0; Mismatches   20; Indels   5; Gaps   4;

Pred. No. 1e-179;

```
Qy   223 caacttccagcagagacgtgcaggagtacagaggagggagtgcgcaggacatgtcagaggacgcacagtaccacaccagc 282
         |||||||||||||||||||||||||||||||||||||||||||||||||||          |||||||||||||||||||
Db     1 CAACTTCCAGCAGAGACGTGCAGGAGTACAGAGGAGGGAGTGCGCAGGACGACGCCCCAGC 60

Qy   283 agagaaagaagtcagtacccagaaatcatctccttggaacagcagggtctgccatcccgat 342
         |||||||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||
Db    61 ACAGAAAAGAAGTCAGTACCCACAGAAATCATCTTCCTTGGAACAGCAGGGTCTGCCATCCCGAT 120

Qy   343 gaagattcgaaatgtcagtgccacacttgtcaacataagcccgacacgtctctgctact 402
         ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   121 GAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACT 180

Qy   403 ggactgtggtgagggcacattgggcagctgtgccgtcaattacggagaccaggtggacag 462
         |||||| |||||||||||| || |||||||||||||||| |||||||||||||||||||||
Db   181 GGACTGTGTGGTGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAG 240

Qy   463 ggtcctgggcaccctggctgtgttgttgtgtcccacctggcacgcagatcacacacgg 522
         |||||||||||||||||||||  ||| ||||||||||| |||||||||||||||||||||
Db   241 GGTCCTGGGCACCCTGGCTGCGTGTTGTTTGTCCCCACCT-GCACGCAGATCACACACGG 299

Qy   523 gcttgccaagtatcttgctgcagagagagaacgcgcctggcatctctttgggaaagccgcttc 582
         |||||||||||||||||||||||||||| |||||||| ||||||||   |||||||||||||||
Db   300 GCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCCTGGCATCTTTGGGAAAGCCGCTTC 359

Qy   583 acccttcgctggtcggttgccccaacagctcaaaagcctggctccagcagtaccacaacc 642
         ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   360 ACCCTTTGCTGGTCGGTTGCCCCAACAGCTCAAAAGCCTGGCTCCAGCAGTACCACAACC 419

Qy   643 agtgccaggaggtccctgcacacatcagtatgtgattcctgccaaatgccttcaggaagggg 702
         ||||||||||||||||||||||| |||||| |||| |||||||||| |||||||||||||||
Db   420 AGTGCCAGGAGGTCCCTGCACCACATCACTGTATGATTCCTGCCAAAATGCCTTCAGGAAGGG 479

Qy   703 ctgagatctccagtcctgcagtggaagaattgaatcagttcgctgtccgaacagtgatt 762
         |||||||||||||||||||||||||||||| |||| ||||||||||   |||||||||||
Db   480 CTGAGATCTCCAGTCCTGCAGTGGAAGAAATTGATCAGTTCGCTGTTGGCAACAGTGATC 539

Qy   763 tggaaagagttcagaccttgtctggtcggcactgcagacgcaagatcggtttggctcgcgctgg 822
         ||||| ||||||||||||| |||| ||||| ||||||| ||| ||||||||| ||||| ||||||
Db   540 TGGAAGAGTTTCAGACCTGTCGTTCGGTGCGCACTGCAGCCATCGGATTGGCTGTTGCGGCTGG 599

Qy   823 tgcacactcctggctcggaaagtggtcattccgggacaccatgccctgcgcaggggctctgg 882
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   823 TGCACACTCCTGGCTCGGAAAGTGGTCATTCCGGGACACCATGCCCTGCGCAGGGGCTCTGG 882
```

```
Db  600  TCGACACCTTCTGGCTGGAAGTCGTCTATTCCGGGACCAATGCCCTTGCGAGGCTCTGG  659
         ||||||||| |||||||||||||| ||||||||||||
Qy  883  tccggatggggaaagatgccaccctcctgatacatgaagccaccctggaagatggtttgg  942
         ||||||||| ||||||||||  |||||||||||||| || |||||||||||
Db  660  TCCGGATGGGGAACGATGCCA-CCTCCTGATTCATGAAGC--ACCTGAAGAGGTACGG  716
Qy  943  aagaggaagcagt-ggaaagacaacagtcccaagcaccaagcgtgggg  998
         |||||||||| || |||||||| ||||||| |||| ||||| |||
Db  717  AAGAGGAAGCACTGGGACACAAGAGACCACGCCCGCGCAAGCGTCCCAAGGCATCAGCGCGGG  773
```

RESULT 2

```
AK004136
LOCUS       AK004136     1510 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus 18 days embryo cDNA, RIKEN full-length enriched
            library, clone:1110037C17, full insert sequence.
ACCESSION   AK004136
VERSION     AK004136.1  GI:12835201
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1510)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,R., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGGCGGCCGCAACTCAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0. Second strand cDNA
            was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAGGATTCAAAGACGCTCAATTAATTAAATTAAAACCCCCCCC 3'], cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5'-end: SstI; 3'-end:
            XhoI. Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1510
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /clone="1110037C17"
                     /dev_stage="18 days embryo"
     CDS             <1..1308
                     /note="putative"
                     /codon_start=1
                     /protein_id="BAB23185.1"
                     /db_xref="GI:12835202"
                     /translation="SSFYSKEEGSTLSVPTVRGECLLKYOLRPKREWORDTTLDCNTD
                     EFIAEALELPSFOESVEEYRKNVOENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSST
                     LVNLSPDKSVLLDCGEGTFGOLCRHYGQOIDRVLCSLIFAVFVSHLHADHHTGILNILL
                     QREHALASLGKPFOPLLVVAPTOLRAWLOOYHNICOEILHHVSMIPAXCLOKGAEVSN
                     TTLERLISLLEETCLEEFOTCLVRHCKHAFGCALVHSSGWKVVYSGDFMPCEALVOM
                     GKDATLLIHEATLEDGLEEEAVEKTHSTTSOAINVGMRMNAEFIMLNHFSORYAKIPL
                     FSPDFNEKVGIAFDHMKYCFGDFPTVPKLPIPDLKALFAGDIEEMVERREKRELRVRA
                     ALLTOQADSPEDREPQQKRAHTDEPHSPQGSKKRSVANTLGARV"
     polyA_signal    1488..1493
                     /note="putative"
     polyA_site      1510
                     /note="putative"
BASE COUNT      373 a    412 c    414 g    311 t
ORIGIN
```

```
Query Match       67.5%;  Score 675.8;  DB 192;  Length 1510;
Best Local Similarity   83.2%;
Matches 781;  Conservative  0;  Mismatches 157;  Indels  1;  Gaps  1;
                         Pred. No. 2.2e-174;
```

```
Qy   63  ccagttccgctgtaagaggaggagggcccaccctcagtgtgccatggttcagggtgaat  122
         |||| |||| ||| |||| |||||||||||||||||| |||||| ||| |||| |||
Db    2  CCAGCTTCTTATAGTAAGGAGGAGAGGGTCCACCCTCAGCGTGCCAACACGTTCGGGGTGAAT  61

Qy  123  gcctcctccaagtaccagtctccgtcccaggagggagggagtggcagaggggatgccattactt  182
         |||||||| ||||| ||| |||||| ||||| ||||||||| |||||||||| |||| |||
Db   62  GCCTCCTTCAAGTATCAGCTCTCCGCCCCCAGAGAGAGCTGGCAGAGGGGATACCACACTCGGACT  121

Qy  183  gcaatcctgaggaaattcatcagttcgaaggcgctgcagcttcccaacttcagcagagcgtgc  242
         ||||  |||| |||||| |||| |||| | |||||||| ||||||||| ||||| |||| |||
Db  122  GCAATACTGATGAATTCAATCAGTCAGGCCCTTGGAGCTCCCAGTTTCAGGAGAGTGTGG  181

Qy  243  aggagtacagggagtgccagaacggccagcgccccagccagaagaaagtcagtacc  302
         ||||| |||| |||| |||| || |||||||||||||| |||||||||| ||||
Db  182  AGGAGTATCGGAGAACGTGCCAGCAAAAACCCAGCCCAGGAGGAGAGCCAGTATC  241

Qy  303  cagaaatcatctccttggaacagggttctgccatccccgatgaagatgaagatgtcagtg  362
         | |||| |||||| |||| |||||| |||| ||||| | |||| ||||| | |||||||
Db  242  CTGAAAATGTCTTCCTGGGTACGGGGTCGCCATCCAATCCAATGAACGACATCCGAAAGTCAGTT  301

Qy  363  ccacacttgtcaacataagcccgacacgtctctgctactggactgtggtgaggcacat  422
         ||||| ||||||||| |||| |||||| |||| | || ||| | ||||| | ||||
Db  302  CCCACACTCGTCAACCTTAGCCCGACAAGTCACAGTCAGTTGCTCTGGATTGTGGAGAGGCACTT  361
```

( )

```
RESULT  3

LOCUS       BE902696      735 bp    mRNA           EST       29-SEP-2000
DEFINITION  601677393F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
            mRNA sequence.
ACCESSION   BE902696
VERSION     BE902696.1  GI:10393148
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 735)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM841 row: j column: 23
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1..735
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3959926"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming;
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT   168 a   206 c   203 g   157 t    1 others
ORIGIN


Query Match         64.4%;  Score 644.6;  DB 141;  Length 735;
  Best Local Similarity  97.3%;   Pred. No. 6.3e-166;
  Matches  719;  Conservative  0;  Mismatches  14;  Indels  6;  Gaps  6;
```

```
Qy    51 tccccctgctcacacaagttccgtctgtaagaaagaggagggcccaccctcagtgtgccatgg 110
         |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     1 TCCCCCTGCTCACACAAGTTTCCGCGTGTAAGAAGGAGGGCCCACCCTCAGTGTGCCATGG   60

Qy   111 ttcaggtgaatgcctcctcaagtaccaagctcccgtcccaggaggaggtggcaggaggatg 170
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TTCAGGATGAATGCCTCCTCAAGTACCAAGCTCCCGTCCCAGGAGGAGGTGGCAGGAGGATG 120

Qy   171 ccattattactt gcaatcctgaggagaattcatagttgaggcgttgcagcttcccaactc 229
         |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCATTATTACTTNGCAATCCTGAGGAGAATTCATAGTTGAGGCGTTGCAGCTTCCCAACTTC 180

Qy   230 cagcgagcgtgcgaggtacaaggaggaggtgcgcagcagggccccagcccagcagagaaa  289
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGCGAGCGTGCGAGGTACAGGAGGAGGTGCGCGGCAGGGCCCAGCCCAGCAGAGAAA    240

Qy   290 agaagtcagtacccagaaaatcatcttccttggaacagggtctgcatccgatgaaagatt  349
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGAAGTCAGTACCCAGAAAATCATCTTCCTTGGAACAGGGTCTGCATCCGATGAAGAATT  300

Qy   350 cgaaatgtcagtgccacacttgtcaacataagccccgacacgtctctgtctactggactgt 409
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGTCTACTGGACTGT 360

Qy   410 ggtggagggcacattggcagctgtgcccgtcattacggactgaccacacagggtttgcc  469
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGTGGAGGGCACATTTGGGCAGCTGTGCCCGTCATTACGGACTGACCACACAGGGTCCTG 420

Qy   470 ggcaccctggctgctgtttgtcccaccctggacgagatcaccacacacggggttgcc  529
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGCACCCTGGCTGCTGTG TTGTGTCCCACCT GCACGCAGATCACCACACGGGGCTTGCC 478

Qy   530 aagtatcttgctgcagagagaacgcgcctgcatcttgggaaagccgcttcacccttt  589
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479 AAGTATCTTGCTGCAGAGAGAACGCGCCCTTGGGCAANC TTGGGAAAGCCGCTTCACCCTTT 537

Qy   590 gctggtggttgccccaaccagctccagaccagtaccacaacagtgcca  649
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   538 GCTGGTGGTTGCCCCAACCAGCTCCAGACCAGTACCACAACAGTGCCA 597

Qy   650 ggaggtcctgcaccacatcagtatgatccttggcaaatgcctcaggagatg 709
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   598 GGAGGTCCTGCACCACATCAGTATGATCCTTGGCAAATGCCTCAGGAGAT  656

Qy   710 ctccagtcctgcagtggaagattgatcagttgctgtggcaacatgtgatt-ggaag  768
         |||||||||||||||||||||||||||||||||||||||||| |||||||
Db   657 CTCCAGTCCTGCAGTGGAAGATTGATCAGTTGCTGTGGCGACATGTGATTTCGGAAG  716

Qy   769 agtttcagacctgtctggt 787
         |||||||||||||||||||
Db   717 AGTTTCAGAACTGTCTGGT 735
```

```
RESULT  4
```

```
Qy   423 ttgggacgctagccgtcattacggaggaccaaggtggacaggtgttcctgggcaccctggctg 482
         ||||||||||||||||||||||||| | ||||||| | |||  |||||||||||||||||
Db   362 TTGGGCAGTTGTGCCGTCATTACGGACTGACCACACAGGGTCCTGTTATGCAGCCTCACGG 421

Qy   483 ctgtgtttgtgtccaccctggacgcagagatcaccacagtatctcttgctg 542
         |||| |||||||||||||||||||| | |||||||||||| ||||||||||
Db   422 CTGTCTTTGTTGTCTCCACCCT-GCACGCCGAGATCACCACACGGGCTTTGCTG 480

Qy   543 cagagagaacgcgcctggcatcttgggaaagccgcttcaccctttgctggttgtt 602
         ||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db   481 CAGAGAGACGCATCCGGTTGGCATCTCTGGGAAACCCTTTGCTTGTTGGTGGTT 540

Qy   603 cccaaccagctcaaagcctggctccagcagccaacaacagtgccaggaggtcctgcac 662
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCCAACCAGCTCAGGGCTCGCAGCCCACACAGTACCCCAACAACCAGTGCCAGGAGGTCCTGCAC 600

Qy   663 cacatcagtatgatcctgaaatgcctcagcagtatcacaacactgccaggagtcctgcca 722
         |||||||||||||||| ||||||||||||||||||| | |||||||||||||||||||
Db   601 CACGTCAGTATGATTGATCAGTTGCTGTGGCGACATGTTGTAAGAGAGTCTCCAATACTACA 660

Qy   723 gtggaaagattgatcagttctgtctgttgcgaacatgtgattggaagagtttcagacctgt 782
         ||||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   661 GTGGAAAGGCTGATCAAGCTTGTGCGTTGGTGGAAACATGTGACTTGAGAACAATTTCAGACCTGC 720

Qy   783 ctgggtcggcaactgcaagacagtcgtttggcgtggtgtgtcacaacctctggctggaa 842
         |||||| |||||||||||||||| |||| |||||||| | |||||||||||||||||
Db   721 CTGGTACGGCGACTGCAAGCANTGCTTTTGGCCTGTGCTGGTACATTCATCTGGCTGGAAA 780

Qy   843 gtggtcttccggggacacatgcgcctctggtccggatgtggggaaagatgcc 902
         |||||| ||||| |||||||||| |||||||||||||||||||||||||||
Db   781 GTCGGTCACTTCGGCGGGGATACCATGCCCTGTCGGGTTCTCAGATNGGGGAAAGANCGCC 840

Qy   903 acctcctgatacatgaagcgaccaccctggaaagatgttttgggagagcagtggaaag 962
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ACCCTCTGATACATCTCAAAGCCCACTCTGCAGCCACCTGGAAAGAGCACTAGAGAAG 900

Qy   963 acacacagcacaacgtcccaaagccatcacagcagtggggatg 1001
         |||||||||||||||||||||||||||||||||||||||||
Db   901 ACACACAGCACAACCACCTCCCAGGTCATTAATGTGGGGATG 939
```

```
LOCUS       BF525432              947 bp    mRNA            EST        11-DEC-2000
DEFINITION  602069517F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4212365
            5', mRNA sequence.
ACCESSION   BF525432
VERSION     BF525432.1  GI:11612793
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9783  row: a  column: 06
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NCI_CGAP_Brn64"
                     /clone="IMAGE:4212365"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.57 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      224 a    257 c    289 g    177 t
ORIGIN
```

```
Query Match
Best Local Similarity    96.5%;
Matches   737;  Conservative    0;
                                      Score 636.8;  DB 150;  Length 947;
                                      Pred. No. 9.4e-164;
                                      Mismatches   17;  Indels   10;  Gaps   8;

Qy     241 gcaggagtacaggaggtgcgacagcgccagccagcagagaagaaagtcagta 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCAGGAGTACAGGAGGAGTGCGCCAGGCACGGCCCCAGCAGAGAAGAAGTCAGTA 60

Qy     301 cccagaaatcatcttccttggaacacaggggtctgccatcccgatgaagattcgaaatgtcag 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCCAGAAATCAACTTTCCTTGGAACACAGGGTCTGCCATCCGATGAAGAATTCGAAATGTCAG 120

Qy     361 tgccacaacttgtcaaacataagcccgacaacgtctctgctactggacgtgtggtgaggcac 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGCCACAACTTGTCAACATAAGCCCCGACAACGTTCTTCTACTGGACTGTGGTGAGGCCAC 180

Qy     421 atttgggcagctgtgccgtcattacggagaccaagggtcctgggcaccctggc 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATTTGGGCAGCTGTGCCGTCATTACGGAGACCAAGGGTCCTGGGCACCCTGGC 240

Qy     481 tgctgtgtttgtgttccaccttggcacgcacgatcaccaacacggttgcgcaagtatctgc 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGCTGTG-TTGTGTCCCACCT-GCACCGACGATCACCAACACGGTGTTGCCAAGTATCTTGC 298

Qy     541 tgcagagaacgcgccttggcatcttggaaagcgcttcaccctttgcttggttg 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 TGCAGAGAACGCGCCTTGGCATCTTGGAAAGCGCTTCACCCTTTGCTTGGTTG 358

Qy     601 cccccaaccagctcaaagcctggctccagcagtaccacacaacagtgccaggaggtcctgc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 CCCCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACACAACAGTGCCAGGAGGTCCTGC 418
```

```
LOCUS       BE795820              761 bp    mRNA            EST        20-SEP-2000
DEFINITION  601590856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
            mRNA sequence.
ACCESSION   BE795820
VERSION     BE795820.1  GI:10217018
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM802  row: P  column: 14
            High quality sequence stop: 761.
FEATURES             Location/Qualifiers
     source          1..761
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NIH_MGC_7"
                     /clone="IMAGE:3945085"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      172 a    206 c    219 g    164 t
ORIGIN
```

```
RESULT  5
BE795820

Qy     661 accacaatcagtatgatcctgccaaatgcctcaggaaggggcttgagaatccagtccctg 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 ACCACAATCAGTATGATGATTCCTGCCAAAATGCCTTCAGGAAGGGGCTGAGAATCCAGTCCCTG 478

Qy     721 cagtggaaagagattgatcagttcgctgtttgcgaacatgttgatttggaagagttcagaact 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 CAGTGGAAAGAGATTGATCAGTTCGCTGTTTGCGAACATGTTGATTTGGAAGAGTTCAGAACT 537

Qy     781 gtctggttgcggcactgcaagcagcgtgtttggctttgcgtctggtgcacacctctggctgaa 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     538 GTCTGGTTGCGGCACTGCAAGCAGCGTGTTTGGCTGTGCGGCT-GTGCACACCTCTGGCTGGA 595

Qy     841 aagtggtctattccggggacaccatcgccctgcgaggcttctggttcgga---tggggaaag 897
           |||||||||||||||||||||||||||||||||||||||||||||||   |||||||
Db     596 AAGTGGTCTATTCCGGGGACACCATCGCCCTGCGAGGCTTCTGGTTCGGACTGGGGAAAGAG 655

Qy     898 atgcacacctcctgatacatgaacgccacacctgaagatgtgttggaaagggaagcagtgg 957
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     656 ATGCCAGCCTCCTGATACATGAACGCCACACCCTGCGAGATGG-TTGGAAGAGGAAGCAGTGG 714

Qy     958 aaaagacacaacagcacacaacgtcccaaagccatcaagcagtgtggggatg 1001
           |||  | |||||||||||||||||||| |||| |||||||
Db     715 AAAGACCACCAGCTCAACAA-GTCCCAAGCATCACAACGGGGGGGGGATG 757
```

```
Query Match    Best Local Similarity  63.6%;     Score 636.4;  DB 140;  Length 761;
Matches  659;  Conservative  0;  Pred. No. 1.1e-163;  Mismatches  1;  Indels  2;  Gaps  2;

Qy   340 gatgaagattcgaaatgtcagtgccacacttgtcaacataagccccgacacgtctctgct 399
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCT 60

Qy   400 actggacctgttgtgagggcacattgggcagctgtccgtgcatcattacgtgacggtgga 459
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACTGGACCTGTTGTGAGGGCACACTTGGGCAGCTGTCCGTGCATCATTACGTGACGGTGGA 120

Qy   460 cagggtcctggggcaccctggctgctgtgtttgtctcccacctggcacgcagagataccaca 519
         |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   121 CAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTTCCCACCT-GCACGCAGATCACCACA 179

Qy   520 cgggcttgccaagtatcttgctgcagagagagaacgcgccttggcatcttcttgggaagccgc 579
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180 CGGGCTTGCCAAGTATCTTGCTGCAGAGAGAGAACGCGCCTTGGCATCTTCTTGGGAAGCCGC 239

Qy   580 ttcacccttgctggtggttgcccccaacaagctggctccccacagcagtaccaca 639
         |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   240 TTCACCCTTGCTGGTGGTTGCCCCCAACAAGCCTGGCTCCCCAGCAGTACCACA 299

Qy   640 accagtgccaggaggtcctgcacacaatcagtatgatgattcctgccaaatgcctcaggaag 699
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   300 ACCAGTGCCAGGAGGTCCTGCACACAATCAGTATGATGATTCCTGCCAAATGCCTTCAGGAAG 359

Qy   700 gggctgagatctccagtcgcagtggaagaggaattgaatcagttcgctgttgcgaacatgtg 759
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   360 GGGCTGAGATCTCCAGTCGCAGTGGAAGAGGAATTGAATCAGTTCGCTGTTGCGGAACATGTG 419

Qy   760 attggaaagagttcagacctgtctgtgtcgcggcacatcgcgttgcgtgtggctgtgtcgc 819
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   420 A-TTGGAAAGAGTTTCAGACCTGTCTGTGTCGCGGCACATCGCGTTGCGTTGGCTGTGTCGC 478

Qy   820 tggtgcacacctctggctggaaagtggtctattccgggacaccaccatgccctgcaggctc 879
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479 TGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGACACCACCATGCCCTGCGAGGCTC 538

Qy   880 tggtccggatggaggaaagatgccacaccctcctgatacatgaagccaccatggtt 939
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 TGGTCCGGATGGAGGAAAGATGCCACACCCTCCTGATACATGAAGCCACCATGGTT 598

Qy   940 tggaagaggaagcagtggaaaaagaccacacagcacacaacgtccaaagccatcagcgtgggga 999
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   599 TGGAAGAGGAAGCAGTGGAAAAAGACCACACAGCACACAACGTCCCAAGCCATCAGCGTGGGGA 658

Qy  1000 tg 1001
         ||
Db   659 TG 660
```

```
Query Match    Best Local Similarity  61.7%;     Score 618;  DB 165;  Length 676;
Matches  629;  Conservative  0;  Pred. No. 1.2e-158;  Mismatches  0;  Indels  1;  Gaps  1;

Qy     1 tcggcagcccacaagattcaaacccagctcaaacttcatccaccggacatctccccctgct 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    48 TCGGCAGCCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCGGACACGTCTCCCCTGCT 107

Qy    61 caccagttccgctgtaaggaagagggcccccaccctcagtgtgcccatggttcagggtga 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   108 CACCAGTTTCCGCTGTAAGGAAGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGA 167

Qy   121 atgcctcccaagtaccagctccgtccacggagggagtggcaggggggatggccattatac 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   168 ATGCCTCCTCCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGGATGGCCATTATAC 227

Qy   181 ttgcaatcctgaggaattcatagttgaggcgctgcagcgttcccaacttccacagcagcgt 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   228 TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCGTTCCCAACTTCCACAGCAGCGT 287

Qy   241 gcaggagtacagaggagtgcgcaggagggaaggcccagcagcagagagagaagaagtcagta 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   288 GCAGGAGTACAGAGGAGTGCGCAGGAGGACGGCCCAGCAGCAGAGAGAAGAAGTCAGTA 347

Qy   301 cccagaaatcatctccttggaacaagggtcctgccatcccgatgaagattcgaaatgtcag 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   348 CCCAGAAATCATCTTCCTTGGAACAAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAG 407

Qy   361 tgccacacttgtcaacataagccccgacacgtctctgctactggacctgttgtgagggcac 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   408 TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTTGGTGAGGGCAC 467

Qy   421 atttgggcagctgtccgtgcatcattacgtgacggtggacagggtcctggggcaccctggc 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   468 ATTTGGGCAGCTGTCCGTGCATCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGC 527

Qy   481 tgctgtgtttgtctcccacctggcacgcagagataccacacgggcttgccaagtatcttgc 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   528 TGCTGTGTTTGTGTTCCCACCT-GCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGC 586

Qy   541 tgcagagagaacgcgccttggcatcttcttgggaagccgcttcacccttgctggtggttg 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   587 TGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTG 646

Qy   601 ccccaaccagctcaaagcctggctccagc 630
         |||||||||||||||||||||||||||||
Db   647 CCCCCAACCAGCTCAAAGCCTGGCTCCAGC 676
```

```
RESULT   7
BE382353
BE382353

LOCUS       BE382353       692 bp     mRNA            EST        21-JUL-2000
DEFINITION  601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
            mRNA sequence.
ACCESSION   BE382353
VERSION     BE382353.1  GI:9327718
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM314 row: o column: 13
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629028"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      152 a    193 c    197 g    150 t
ORIGIN

Query Match             59.9%;  Score 599.8;  DB 166;  Length 692;
Best Local Similarity   98.5%;
Matches  637;  Conservative   0;  Mismatches  7;  Indels   3;  Gaps   3;

Qy      336 tccggatggaagattcgaaatgtcaacatacaagacccgacacgtctc 395
            ||||||||||||||||||||||||||||||||||||||||||| |||||
Db        1 TCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCGACACGTCTC   60

Qy      396 tgctactggactgtgttgaggtgcacatttgggcacagcaacaccaggaccagg 455
            ||||||||||||||| |||||||||||||||||| || |||||||||||||||||
Db       61 TGCTACTGGACTGTGGTTGGGCACCAGCAACAACGCCACTGTCTGCCTCGCCTGTTACGGACGAG 120

Qy      456 tggacaggtcctggacaccctggctgctgtgtttgtgtccaacctgacagaatac 515
            |||||| ||||||||||||||||||||||| || |||||||||||||||||||||||
Db      121 TGGACAGGGTCCTGGGCACCCTGGCTGCTGCGTGTTTGTGTCCAACATAAGCCCGACAGATCAC 179

Qy      516 caacacggcttgcaagtatcttgctgcagagaaacgcgccttggcatcttggggaag 575
            ||||||||||||||||||||||||| || ||||| || ||||| || |||||||||||
Db      180 CACACGGCCTTGCCAAGTATCTTGCTGCGAGAGAACCGCGCCTTGGCATCTTTGGGAAAG 239

Qy      576 ccgctcaccctttgcttggtgttgccccaccagctcaaagcttcagcagtac 635
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      240 CCGCTTCACCCTTTGCTTGGTGGTTGCCCCAACCAGCTCAAAAGCCTCAGCCAGCAGTAC 299

Qy      636 cacaaccagtgcaggagggtcctgcacacacatcagtatgatttcctgccaaatgcctttcag 695
            |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
```

```
RESULT   8
BG519751

LOCUS       BG519751       884 bp     mRNA            EST        02-APR-2001
DEFINITION  6025789O2F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
            mRNA sequence.
ACCESSION   BG519751
VERSION     BG519751.1  GI:13515513
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM176 row: d column: 01
            High quality sequence stop: 859.
FEATURES             Location/Qualifiers
     source          1..884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503184"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      199 a    251 c    249 g    185 t
ORIGIN

Query Match             57.6%;  Score 576.4;  DB 154;  Length 884;
Best Local Similarity   96.3%;
Matches  622;  Conservative   0;  Mismatches  21;  Indels   3;  Gaps   3;
```

```
Db      300 CACAACCAGTGCAGCCAGGAGGTGCCTGACCCACCACATCAGTATGATATGATTCCTGCCAAATGCCTTCAG 359

Qy      696 gaagggggctgaagatctccagtcctgcagtggaaagattgatcagtttcgctgttgcgaaca 755
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      360 GAAGGGGGCTGAAGATCTCCAGTCCTGCCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACA 419

Qy,     756 tgtgattttggaagagatttcagaaccgttctgtcggcgcactgcaagacatggcttggctgt 815
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      420 TGTGATTTTGGAAGAGTTTTCAGACCTTGTCTGTGGCGCCACTGCAAGACATGGCTTGGCTGT 479

Qy      816 gcgctggtgcacacctctggctggaaagtggttctattccgggggacacctgccctgcgag 875
            |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      480 GCGCTGGTGCACACCTCTGGCTGGAAAGTGGTTCTATTCCGGGGACACCTGCCCTGCCGAG 539

Qy      876 gcctctggtccggatggggaaagtgccaccctcctgataccaatgcccaccctggaaagat 935
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      540 GCCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGATACAATACAGGAAAGCCA-CCTGGGAAGAT 598

Qy      936 ggtttggaagagaagcagtggaaaaagaacacacacagcacacgtccca 982
            ||||||||||||||||||||||| |||||||||||||||||||||||| |
Db      599 GGTTTGGCAGAGAAGGAAGCCGTGG-AAAGACACACAGCACACGTCCCAA 644
```

```
Qy   140 ctccgtcccaggaggaggagtggcaaggagtgccattattacttgcaatcctgaggaattc 199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CGCCAGGCAGGGAGTGGCACCATGGGATAGCCATTGATTACTTGCAATCCTGAGGAATTC 296

Qy   200 atagttgaggcgctgcaagcttcccacttccagcagagccgtgcaggagtacaggaggagt 259
             || | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 ATAGTTGAGGCGCTGCAAGCTTCCCAACTTCCAGCAGAGCCGTGCAGGAGTACAGGAGGAGT 356

Qy   260 gcgcaggacggcgcccagcagcagaaagaagtcagtacccagaaatcaatcttcctt 319
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 GCGCAGGACGGCGCCCAGCAGCAGAAAAGAAGTCAGTACCCAGAAATCANTCTTCCTT 416

Qy   320 ggaacagggtctgccatcccgatgaagattcgaaatgtccagtgccacacttgtcaacata 379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 GGAACAGGGTCTGCCATCCCGATGAAGANTTCGAAANGTCAGTGCCACACTTGTCAACATA 476

Qy   380 agcccgacacagtctctgctactggaggtcgtggaggcacattgggcagctgtgccgt 439
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 AGCCCGACACAGTCTCTGCTACTGGAGGTCGTGGTGAGGGCACATTGGGCAGCTGTGCCGT 536

Qy   440 cattacggagaccacggtggacaagggtcctggcacccctggctcgtctgtgtttgtcccac 499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 CATTACGGAGACCACGGTGGACAAGGGTCCTGGCACCCCTGGCTCGTCTGTGTTTGTGTCCCAC 596

Qy   500 ctggcacgcagatcacacaacgggcttgcaagtatctttgctgcagagagaacgcgcctt 559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   597 CT-GCACGCAGATCACACAACGGGCTTGCAAGTATCTTGCCTGCAGAGAACGCGGCCTT 655

Qy   560 ggcatcttttgggaaaccgcgctcacccttgctggttggtgccccaacacgactcaaagc 619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   560 GGCATCTTTTGGGAAAACCGCTTCACCCTTTCCTGGGTGGTTGGCCCCCAACCAGCTCAAAGC 715

Qy   620 ctggctccagcagtaccacaacaacagtgccagaggtcctgcaagtatcttgctgcagagagaa...tcagtatgatcc 678
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db   656 GGCATCTTTTGGGAAAACCGCTTCACCCTTTCCTGGGTGGTTGGCCCCCAACCAGCTCAAAGC 715

Qy   716 ctGGCTCCAGCAGTACCACAACAACAGTGCCAGAGGTCCTGCACCCCAATTCACTATGATTC 775
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   716 ctGGCTCCAGCAGTACCACAACAACAGTGCCAGAGGTCCTGCACCCCAATTCACTATGATTC 775

Qy   679 ctgccaaatgccttcaggaaggggctgagatctccagtgtcctgcagtggaaagattgatca 738
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   679 CTGGCTCCAGCAGTACCACAACAACAGTGCCAGAGGTCCTGCACCCCAATTCACTATGATTC 775

Qy   776 CTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTGTCCTGCAGTGGAAAGATTGATCA 835
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   776 CTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTGTCCTGCAGTGGAAAGATTGATCA 835

Qy   739 gttccgtcgtgcgaacatgtgatttggaa-gagtttcagaacctgtc 783
             ||||||||||||||||||||||||||||| |||||||||||||
Db   836 GTTCCGTCGTTCGGAACATGTGATTTGGAACACGAGTTTCAGAACCTGTC 881
```

```
Qy   385 cgacacgtctctgctactggaggtcgtggtgaggggcacattgggcagctgtgccgtcatta 444
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGACACGTCTCTGCTACTGGACTGTGGTAGGCACGTGAGGGCACATTTGGGCAGCTGTGCCGTCATTA 60

Qy   445 cggagaccaggtggacaagggtcctggcacccctggctctgtgtttgtgtccccacctggc 504
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CGGAGACCAGGTGGACAGGGTCCTGGCACCCTGGCTCTGTGTTG-TTGTGTCCCCACCT-GC 118

Qy   505 acgcagagatcacacaacgggcttgcaagtatctctgctcagagagagagaacgcgcctggcat 564
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   119 ACGCAGAGATCACACACACGGGCTTGCCAAGTATCTTGCCTGCAGAGAGGAACGCGCCCTTGGCAT 178

Qy   565 ctttggggaaaccgcgttcacccttgtcgttggtggtgccccaaccagctcaaagctggc 624
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   179 CTTTGGGAAAGCCGCTTCACCCTTGTCGGTGGTGCCCCAACCAGCTCAAAGCCTGGC 238

Qy   625 tccagcagtaccacaacaacagtgccagaggtcctgcaccacatcagtatgatcctgcca 684
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   239 TCCAGCAGTACCACAACAACAGTGCCAGAGGTCCTGCCAGGAGTCCTGCACCACATCAGTATGATTCCTGCCA 298

Qy   685 aatgccttcaggaagggggctgagatctccagtgtcctgcagtggaaagattgatcagttcgc 744
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   299 AATGCCTTCAGGAAGGGGGCTGAGATCTCCAGTGTCCTGCACGTGGAAGAGATTGATCAGTTCGC 358

Qy   745 tgttgcgaacatgtgatttggaagagttcagaacctgtctcggttgcggcactgcaagcatg 804
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   359 TGTTGCGAACATGTGATTTGGAAGAGTTTCAGAACCTGTCTCGGTTGCGGCACTGCAAGCATG 418

Qy   805 cgttggcttgctggtgcacacctctggctggaaagtggtctattccggggacaca 864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   419 CGTTGGCTTGCGTGGCCGTGGTGCACACTCTGGCTGGAAAGTGGTCTATTCC-GGGACACCA 477

Qy   865 tgccctgcgaggctctgttcccggtggggaaagtgccacccctcctgacatacatgaagcca 924
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   478 TGCCCTGCGCAGGCTCTGTTCCCGGATGGGAAGATGCCACCCCTCTGATACATGAAGCCA 537

Qy   925 -ccctggaagatggtttggaaaggaggaagcaagtggaaaagacacagcacaaacgtcccaa 983
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   538 CCCCTGGAAGATGGTTTGGAAGAGGAGGAAGCAAGTGGAAAAGACACAGCACAAACGTCCCAA 597

Qy   984 gccatcaagcgtgggatg 1001
             ||||||||||||||||||
Db   598 GCCATCAAGCGTGGGGATG 615
```

Best Local Similarity   56.8%;   Score 568.4;   DB 141;   Length 745;
Matches 613;   Conservative   0;   Mismatches   1;   Indels   4;   Gaps   4;
Query Match   99.2%;   Pred. No. 5e-145;

ACCESSION   BG335963
VERSION   BG335963.1 GI:13142401
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 614)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1222 row: o column: 02
            High quality sequence stop: 614.
FEATURES   Location/Qualifiers
     source   1..614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4542553"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: Placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT   135 a   169 c   178 g   132 t
ORIGIN

Query Match   56.7%;   Score 567.6; DB 152; Length 614;
Best Local Similarity   97.4%;   Pred. No. 7.8e-145;
Matches 598; Conservative   0; Mismatches 14; Indels 2; Gaps 2;

Qy   366 cacttgtcaacataagcccgaccacgtctctgctactggactgtggtgaggggcacattg 425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   2 CACTTGTCAACATAAGCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGGCACATTTG 61

Qy   426 ggcagctgtgccgtgtcattacgtgaagaccaggtgtcctgggcacctggctgctg 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GGCAGCTGTGCCGTGTCATTACGTGAAGAGGTGTCCGGGCACCCTGGCTGCTG 121

Qy   486 tgtttggtcccacctggcacgcagcagatcaccacacgggcgttgccaagtatcttgctgag 545
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TGTTTGGTCCCACCT-GCACGCAGATCACCACACGGGCGTTGCCAAGTATCTTCGCTGCAG 180

Qy   546 agggaaccgcgccttggcatcttgggaaagccgcgccgttcaccctttgctcgtcgttggttgccccc 605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGAGAACCGCGCCTTGGCANCTTTGGGAAAGCCCGCTTCACCCTTTGCTCGTGTGGTTGCCCCC 240

Qy   606 aaccagctcaaagctgcctggctccagcagtaccacaccagtgccaggagggtcctgcaccac 665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AACCAGCTCAAAGCCTGGCTCCAGCAGTACCACCAGTGCCAGGAGGTCCTGCACCAC 300

Qy   666 atcaagtatgatgattcctgccaaatgcctttcaggagggctgagatctccagtcctgcagtg 725
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATCAGTATGATGATTCCTGCCAAATGCCTTCAGGGGAGATCTCCAGTCCTGCACTTG 360

Qy   726 gaaagattgatcagttcgctgttgcgaacaatgtgattggaaggtttcagacactgtctg 785
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAAAGATTGATCAGTTCGCTGTTGCGAACAATGTGATTTGGAAGAGTTTCAGACACTGTCTG 420

RESULT 11
BE383336   BE383336   664 bp   mRNA   EST   21-JUL-2000
LOCUS   6012982495F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
DEFINITION mRNA sequence.
ACCESSION   BE383336
VERSION   BE383336.1 GI:9328701
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 664)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM313 row: a column: 13
            High quality sequence stop: 662.
FEATURES   Location/Qualifiers
     source   1..664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3628308"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   154 a   199 c   174 g   137 t
ORIGIN

Qy   786 gtgcgcgcactgcaagcgatgcgtttcggctcgtgtcgctggtgcacactctggctggaaagtg 845
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTGCGCGCACTGCCAACACATGCGTTTGGCTGTGCGCGTGGTGCACACTCTGGCTGGAAAGTG 480

Qy   846 gtctattccgggacaccatgcccttgcgaggtctggttccggatgggg-aaagatgtccac 904
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GTCTATTCCGGGACACCATGCCCTTGCGAGGTCTGGTTCCGGATGGGGAAAGATGCCAC 540

Qy   905 cctcctgatacatgaagccaccctgggaaagtggttttggaggaagcagagtgaaagac 964
          |||||||||||||||||||||||||||||||||||||||| | | ||||||||||||||
Db   541 CCTCCTGATACATGAAGCCACCCTTGGGAAAGTGGTTTGGAAGATGGTTGGAAGCAGCAGTGAAAGAC 600

Qy   965 acacagcacaacgt 978
          ||||||||||||||
Db   601 ACACAGCACAACGT 614

BASE COUNT   135 a   169 c   178 g   132 t
ORIGIN

Query Match   56.4%;   Score 564.8; DB 166; Length 664;
Best Local Similarity   99.2%;   Pred. No. 4.7e-144;
Matches 599; Conservative   0; Mismatches 2; Indels 3; Gaps 3;

Qy   1 tcgcagcacaagattcaaaccagctcaacctcatccaccaggacaactctccccctgct 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48 TCGCAGCCACAAGATTCAAACCCAGTCAACCTCATCCACCAGGACATCTTCCCCCTGCT 107

Qy   61 caccagtttccgcctgtaagaggaggcgcccacctcagtgtgcccatcggttcaggtga 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   108 ...

```
Db  108 CACCAGTTTCCGCTGTGTAAGGAGGAGGCGCCCCACCCTCAGTGTGTCCCATGGTTCAGGGTGA 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 atgcctccttcaagtaccagctccgtccccaggagggggatggatgtcaattattac 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  168 ATGCCTCCTTCAAGTACCAGCTCCGTCCCCAGGAGGGAGGGATGGATGTCCAATTATTAC 227
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181 ttgcaatccttcgaggaattcatagttgaggcgctgcagcttcccaacttccagcagaggt 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  228 TTGCAATCCTTCGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGACCGT 287
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  241 gcaggagtacagcaggagtgcgcaggaacggcccagccccagcagcagaggaaaaagagtcagta 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  288 GCAGGAGTACAGCAGGAGTGCCCAGGACGGCCCAGCCCCAGCAGAGGAGAAAGAGTCAGTA 347
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  301 cccagaaatcatcttccttggaacaggtctgccacatcccgatccccagtgaaatgtcag 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  348 CCCAGAAATCATCTTCCTTGGAACAGGTCTGCCACATCCCGATGAAGATTCGAAATGTCAG 407
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  361 tgccacacttgtcaacataagcccgacacgtctctgctactggactgtggttgaggcac 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  408 TGCCACACTTGTCAACATAAGCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCAC 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  421 atttgggcagctgtccgtcattacggagacacaggtggacagggtcctgggacccctggc 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  468 ATTTGGGCAGCTGTCCGTCATTACGGAGACCAGGGTGGACAGGGTCCTGGGCACCCTGGC 527
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  481 tgctgtgtttgtgtcccacctggcacgcagagatcacacacacgggcttgccaagtatctggc 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  528 TGCTGTGTTTGTGTCCCACCT-GCACCAGATCACACACACGGGCTTGCCAAGTATCTTGC 586
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  541 tgc-agagagaacgcgccctggccatctcttgggaaagccgcttcacccctttgctggttggtt 599
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  587 TGCAAGAGAGAACGCGCCCTTGGCATCTTTGGGAAAG-CGCTTCACCCTTTGCTGGTGTG 645
        ||||
Qy  600 gccc 603
        ||||
Db  646 cccc 649
```

RESULT 12
LOCUS   BE537086   713 bp   mRNA   EST   09-AUG-2000
DEFINITION   601063325F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3449742 5',
           mRNA sequence.
ACCESSION   BE537086
VERSION   BE537086.1   GI:9765731
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 713)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: Incyte Genomics, Inc.
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM8427 row: a column: 07
           High quality sequence stop: 710.
FEATURES   Location/Qualifiers
     source   1..713
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3449742"
                      /clone_lib="NIH_MGC_10"
                      /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                      Average insert size 1.5 kb. Library prepared by Life
                      Technologies."
                      /lab_host="DH10B"
                      /cell_line="MGC36"
BASE COUNT   166 a   187 c   204 g   155 t   1 others
ORIGIN

Query Match   56.2%;   Score 562.6;   DB 136;   Length 713;
Best Local Similarity   92.9%;   Pred. No. 1.9e-143;
Matches 659;   Conservative   0;   Mismatches   34;   Indels   16;   Gaps   6;

```
Qy  116 ggtgaatgcctcctcaagtaccagctccgtcccagcaggaggaggagtggcaggaggggatggccatt 175
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8 GGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGCAGGAGGAGGAGTGGCAGGATGCCATT 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  176 attacttgcaatcctgaggaaattcaatagttgaggcgcgctgcagcttcccaacttccagcag 235
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 ATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCCTGCAGCTTCCCAACTTCCAGCAG 127
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  236 agcgtgcaggagtacaggagtggcgcaggaacggcccagccccagccccagcagagagaagt 295
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 AGCGTGCAGGAGTACAGGAGTGCCGCAGGC--AGCCCAGCAGAGGAGAAAAGAAGT 185
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  296 cagtacccagaaatcatctccttggaacaggttctgccatcccgatgaagattcgaaat 355
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  186 CAGTACCCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAAT 245
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  356 gtcagtgccaacacttgtcaacataagcccgacacgtctctgctactggactgtggtgtgag 415
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  246 GTCAGTGCCACACTTGTCAACATAAGCCCGACACGTCTCTGCTACTGGACTGTGGTGAG 305
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  416 ggcacaattgggcagctgtgccgtcaattacggagagaccaggtgggaca--gggtcctgggca 473
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  306 GGCACGTTTGGGCAGCTGTGCCGCCGTCATTACGGAGACCAGGTGGACAGGGTGGTCCTGGGGCA 365
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  474 ccctg-gctgctgtgtttgtgtccacccggcacgcagatcaccacacgggcttgccaag 532
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  366 CCCTGTGCTGCTGTGTGTATGTGTCCCACCATGCACCAGATCACCACACACGGGCTTGCCAAG 425
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  533 tatcttgctgcagagagaggaacgcgccctggc-atcttgggaaggcg------cttc 582
        |||||||||||| ||| ||||||||||||||| |||||||||| |     ----
Db  426 TATCTTGCTGCCAGAGAGAGGAACGCGCCTTGGGAAATCTTTGGAAAGCGCNCTTCAACCCTTT 485
        ||||||||||||||||||||||||||||||||||||||||||||||
Qy  583 acccttttgctggttggtttgccccccaacccagctcaaaagcgtccagcagtaccacacaacc 642
        |||  |||  | ||| ||||||||||||||||   ||||||||||||||||||||||||
Db  486 TGCTTGGTGGGTTGGCCCCCCAAAGCCAGGGCTCAAACCCTGGCCTCCAGCAGTACCACACAACC 545
        ||||||||||||||||||||||||||||||||||||||||||||||
Qy  643 agtgccaggggagtgtcctgcaccacatcagtatgatctct 702
        |||||||  | || ||||| ||||||||||||||||||
Db  546 AGTGCCCACGGAGGGTCCTGACCACACCATCAGTATGATT 605
        ||||||||||||||||||||||||||||||||||||||
Qy  703 ctgaggatctccagtcctgcagtgaaagagattgatcagttc 762
        |||||  |||||||||||||||||| |||| |||||  || 
Db  606 CTGAGACATCTCCAGTCCTGCAGTGGAAAAGATGAAATCAGT 665
        ||||||||||||||||||||||||||||||||||||||
Qy  763 tggaagagagtttcagaacctgtctggtgtgccgcgcactcgcaaaagcaatgcgtttgg 811
        ||||||||||||||||||||||||||||||||||||| |||||||||| || 
Db  666 TGGGAGAGAGTTTCAGAACCTGTTCTGGTGGTCGCGG-ACTGCAAGCAATGCGTTTGGG 713
```

RESULT 13
LOCUS   BE619259   823 bp   mRNA   EST   20-OCT-2000
DEFINITION   601473130F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5',
           mRNA sequence.
ACCESSION   BE619259
VERSION   BE619259.1   GI:9890197
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens

REFERENCE 1 (bases 1 to 823)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: DCTD/DTP/Gazdar
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM9636 row: k column: 08
           High quality sequence stop: 695.
FEATURES   Location/Qualifiers
     source   1..823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3876223"
                     /clone_lib="NIH_MGC_68"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."
BASE COUNT   219 a   259 g   160 t
ORIGIN
                                                       185 a

Query Match   54.1%; Score 541.4; DB 138; Length 823;
Best Local Similarity   99.3%; Pred. No. 1.3e-137;
Matches 575; Conservative   0; Mismatches   1; Indels   3; Gaps   3;

Qy   423 ttgggcagctgtgccgtcattacggagaccaaggtggacaaggtcctgggcaccctggctg 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 TTGGGCAGCTGTGCCGTCATTACGGAGACCAAGGTGGACAAGGTCCTGGGCACCCTGGCTA 69

Qy   483 ctgtgtttgtgtcccacctgcaacgcagatcaccacacgggcttgccaagtatcttgctg 542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 CTGTG-TTGTGTCCCACCT-GCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTG 127

Qy   543 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggttgcc 602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128 CAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTGCTGGTTGCC 187

Qy   603 cccaacaagctcaaagcctggctggtcttccagccagtaccacccactgtccaggaggtcctcgac 662
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   188 CCCAACAAGCTCAAAGCCTGGCTGGTCTTCCAGCCAGTACCACCCACTGTCCAGGAGGTCCTCGAC 247

Qy   663 cacatcagtatgattcctgccaaatgccttcaggaaggagggtcagtacctcctgca 722
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGCTGAGATCTCCAGTCCTCA 307

Qy   723 gtggaaagatgatcagttcgtctgtgcgaaacatgtgatttggaaagagtttcagacctgt 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308 GTGGAAAGATGATCAGTTCGTCTGTGCGAAACATGTGATTTGGAAAGAGTTTCAGACCTGT 367

Qy   783 ctggtgcggcaactgcaagacatccgtttggctgtgtgcacacctctggctggaa 842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   368 CTGGTGCGGCACTGCAAGCACATCCGTTTGGCTGTTGCGCTGTGGTCCACACCTCTGGCTGGAAA 427

Qy   843 gtggtctcatctccgggacacccatgcccctgcggaggctctcggctgctggagatgcc 902
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   428 GTGGTCTCATTCCGGGGACACCCATGCCCTGCGGAGGCTCTCGGCTGCTGGAGATGCC 487

Qy   903 accctcctgatcatagaaggccaccccctggaaaagtgttggaagagaggaagcagtggaaag 962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   488 ACCCTCCTGATGATCATAGAAGGCCACCCCCTGGAAGATGGTTTGGAAGAGGAAGCAG-GGAAAAG 546

RESULT 14
Db   547 ACACACAGCACAAACGTCCCAAGCCCATCAGCCGTGGGGATG 585

Qy   963 acacacagcacaaacgtcccaaagccatcaggtgggggatg 1001
          |||||||||||||||||||||||||||||||||||||||||

BF311926
LOCUS   BF311926   915 bp   mRNA   EST   21-NOV-2000
DEFINITION 601897767F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126864 5',
           mRNA sequence.
ACCESSION BF311926
VERSION   BF311926.1 GI:11259697
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 915)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
           Plate: LLCM1017 row: f column: 17
           High quality sequence stop: 672.
FEATURES   Location/Qualifiers
     source   1..915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4126864"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT   206 a   262 c   288 g   159 t
ORIGIN

Query Match   53.5%; Score 535.4; DB 147; Length 915;
Best Local Similarity   98.0%; Pred. No. 5.9e-136;
Matches 595; Conservative   0; Mismatches   6; Indels   6; Gaps   5;

Qy   1 tcgcagccacaagattcaaacccagctcaacctcatccaccgagacaatctctccctgct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 TCGCAGCCACAAGAGTTCAAAACCCAGCTCAACCTCATCCACCACCGGACATCTTCCCCTGCT 127

Qy   61 cacagtttccgtctgtaagagaaggaggcccccacccctcagtgtgccatggttcagggtga 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128 CACCAGTTTCCGTCTGTAAGAAGAGGAGGGCCCCCACCCCTCAGTGTGCCCATGGTTCAGGGTGA 187

Qy   121 atgcctcctcaagtaccagctccgtcccaggaggaagtggcagaggaggatgccattattac 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   188 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGAGGATGCCATTATTAC 247

Qy   181 ttgcaatcctgaggaattcatagttgaggcgctgcagcttcccaactccagcagagcgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248 TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTCCAGCAGAGCGT 307

Qy   241 gcaggagtacagggaggagtcgcaggacacggcccagccccagcagagacagaaagaagtcagta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
LOCUS       BE747163      992 bp    mRNA                  EST          15-SEP-2000
DEFINITION  601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
            mRNA sequence.
ACCESSION   BE747163
VERSION     BE747163.1  GI:10161155
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 992)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM524 row: I column: 14
            High quality sequence stop: 781.
FEATURES             Location/Qualifiers
     source          1..992
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_9"
                     /clone="IMAGE:3838237"
                     /tissue_type="adenocarcinoma cell line"
BASE COUNT      225 a    287 c    313 g    167 t
ORIGIN
```

RESULT 15
BE747163

```
Query Match  50.5%;  Score 505.6;  DB 139;  Length 992;
Best Local Similarity  97.5%;  Pred. No. 9e-128;
Matches 546;  Conservative  0;  Mismatches  9;  Indels  5;  Gaps  3;

Qy      1 tcgcaagccacaagattcaaacccagctcaacctcatccacccggacatctcccctgct   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    148 TCGGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTCCCCTGCT  207

Qy     61 caccagttccgctgtaagaagaggaggggccccaccctcagtgtgtccatggttcagggtga  120
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    208 CACCAGTTTCCGCTGTAAGAAGAGGAGGGGCCCCACCCTCAGTGTGTCCATGGTTCAGGGTGA  267

Qy    121 atgcctcctcaagtaccagctccgtcccagaggagggagtggcagaggatgccattattac  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGATGCCATTATTAC  327

Qy    181 ttgcaatcctgaggaattcatagttgagcgtgcaagcttcccaacttccagcagagcgt  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    328 TTGCAATCCTGAGGAATTCATAGTTGAGCGTGCAAGCTTCCCAACTTCCAGCAGAGCGT  387

Qy    241 gcaggagtacaggaggagtcgcgacagacgcccagccagcagagagaaagaagtcagta  300
          ||||||||||||||||||||||||||||||| ||||||| |||||||||||||||||||
Db    388 GCAGGAGTACAGGAGGAGTCGCGACAGACGG-CCAGCCCAGCAGAGAGAAAGAAGTCAGTA  446

Qy    301 cccagaaatcatcatcttcctctggaacaagggttctgccatcccgatgaagaattcag  360
          ||||||||||||||| ||||||| |||||| |||||||||||||||||||||||||||
Db    447 CCCAGAAATCATCATTCTTCCTTGGAACAGGGTTCTGCCATCCCGATGAAGAATTCAG  506

Qy    361 tgccacacttgtcaacataagccccgacacgtctctgtcactggactgtgggtgagggcac  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    507 TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGTCACTGGACTGTGGGTGAGGGCAC  566

Qy    421 atttgggcagctgtgccgtcattacgggaccacaggtggacaggtgtcctgggcaccctggc  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 ATTTGGGCAGCTGTGCCGTCATTACGGGACCACAGGTGGACAGGTGTCCTGGGCACCCTGGC  626

Qy    481 tgctgtgttttgtgtcccacctggcacgcagatcaccacacgggcttgccaagtatcttgc  540
          ||||||||||||||||||||||| |||||||||  ||||||||||| ||||||||||||||
Db    627 TGCTGTGTTTTGTGTCCCACCTT-GCACGCAGAT--CACACACGGGCTGCAAGTATCTTGC  682

Qy    541 tgcagagagaaacgcccttg  560
          |||||||||||||||| ||
Db    683 TGCAGAGAGAAACGCCCCTGG  702
```

```
Db    308 GCAGGAGTACAGGAGGAGTCGCGACAGACGG·CCAGCCCACACGAGAAAAAGTCAGTA  366
          ||||||||||||||||||||||||||||||| |||||||
Qy    301 cccagaaatcatcttcctctggaacaggttctgccatcccgatgaagaatgtcag  360
          |||||||||||||||||||| |||||
Db    367 CCACAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGAATGTCAG  426

Qy    361 tgccacacttgtcaacataagccccgacaagtctctgtcactgactgtgtgggcac  420
          |||||||||||||||||||||||||||
Db    427 TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGTCACTGGACTGTGGTGGGCAC  486

Qy    421 atttgggcagctgtgccgtcattacgggagaaccaggtgggtcctgggcaccctggc  480
          ||||||||||||||||||||||||||||||
Db    487 ATTTGGGCAGCTGTGCCGTCATTACGGGAGAACCAGGTGGACAGGTGTTCAGGGTGA  546

Qy    481 tgctgtgttttgtgtcccacctggcacgcagatcaccacacgggcttgcc-aagtacttg  539
          ||||||||||||||||||||||| |||||||
Db    547 TGCTGTGTTTTGTGTCCCACCT-GCACGCAGATCACCACACGGGCTTGCCAAAGTATCTTG  605

Qy    540 ctgcagagagaaacgcccttgggaaacgcgttcacccctttgctggttgtt  599
          |||||||||||||
Db    606 CTGCAGAGAGAACGCGCCCTGGG--ATCTTGGGAAAACCGCGCTTCACCC-TTGCTGGTTGGTG  662

Qy    600 gccccca  606
          ||||| |
Db    663 GCCCCAA  669
```

Search completed: October 28, 2001, 21:23:18
Job time: 6051 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: October 28, 2001, 21:48:07 ; Search time 1394.01 Seconds
                                              (without alignments)
                                              11106.963 Million cell updates/sec

Title: SEQIINS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacaagattcaaa.........aagcatcaagcgtgggatg 1001

Scoring table: OLIGO_NUC
                  Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size :   8

Total number of hits satisfying chosen parameters: 2254642

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   GenEmbl1:*
                  1: gb_ba1:*
                  2: gb_ba2:*
                  3: gb_ba3:*
                  4: gb_in1:*
                  5: gb_in2:*
                  6: gb_in3:*
                  7: gb_om:*
                  8: gb_ov:*
                  9: gb_pat1:*
                 10: gb_pat2:*
                 11: gb_ph:*
                 12: gb_pl1:*
                 13: gb_pl2:*
                 14: gb_pl3:*
                 15: gb_pl4:*
                 16: gb_ba1:*
                 17: gb_ba2:*
                 18: em_fun:*
                 19: em_htgo_hum:*
                 20: em_htgo_inv:*
                 21: em_htgo_rod:*
                 22: em_htg_hum1:*
                 23: em_htg_hum2:*
                 24: em_htg_hum3:*
                 25: em_htg_hum4:*
                 26: em_htg_hum5:*
                 27: em_htg_hum6:*
                 28: em_htg_hum7:*
                 29: em_htg_hum8:*
                 30: em_htg_inv1:*
                 31: em_htg_inv2:*
                 32: em_htg_other:*
                 33: em_htg_rod:*
                 34: em_hum1:*
                 35: em_hum2:*
                 36: em_hum3:*
                 37: em_hum4:*
                 38: em_hum5:*
                 39: em_hum6:*
                 40: em_hum7:*
                 41: em_in:*
                 42: em_om:*
                 43: em_or:*
                 44: em_ov:*
                 45: em_pat:*
                 46: em_ph:*
                 47: em_pl:*
                 48: em_ro:*
                 49: em_sts:*
                 50: em_sy:*
                 51: em_un:*
                 52: em_v1:*
                 53: gb_sts1:*
                 54: gb_sts2:*
                 55: gb_sts3:*
                 56: gb_sy:*
                 57: gb_un:*
                 58: gb_vil:*
                 59: gb_vl2:*
                 60: gb_htg1:*
                 61: gb_htg2:*
                 62: gb_htg3:*
                 63: gb_htg4:*
                 64: gb_htg5:*
                 65: gb_htg6:*
                 66: gb_htg7:*
                 67: gb_htg8:*
                 68: gb_htg9:*
                 69: gb_htg10:*
                 70: gb_htg11:*
                 71: gb_htg13:*
                 72: gb_htg14:*
                 73: gb_htg15:*
                 74: gb_htg16:*
                 75: gb_htg17:*
                 76: gb_htg18:*
                 77: gb_htg19:*
                 78: gb_htg20:*
                 79: gb_htg22:*
                 80: gb_htg23:*
                 81: gb_htg24:*
                 82: gb_htg25:*
                 83: gb_pr1:*
                 84: gb_pr2:*
                 85: gb_pr3:*
                 86: gb_pr4:*
                 87: gb_pr5:*
                 88: gb_pr6:*
                 89: gb_pr7:*
                 90: gb_pr8:*
                 91: gb_pr9:*
                 92: gb_rol:*
                 93: gb_ro2:*
                 94: gb_in4:*
                 95: gb_pr10:*
                 96: em_ba3:*

Handwritten note (boxed): Insertion of 1
Oligo search of
1000 residues
marketing
insertion
site.
at least
8 mers

Seq ID NO 1

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	880	87.9	2908	89	AF304370	AF304370 Homo sapi
2	829	82.8	2997	91	BC001939	BC001939 Homo sapi
3	829	82.8	3006	91	BC004158	BC004158 Homo sapi
4	750	74.9	2976	89	AK001392	AK001392 Homo sapi
5	371	37.1	2908	89	AF308698	AF308698 Pan trogl
6	343	34.3	2893	89	AF308694	AF308694 Gorilla g
7	124	12.4	118788	85	AC005277	AC005277 Homo sapi
8	123	12.3	740	89	AF304369	AF304369 Homo sapi

RESULT 1

#					Acc1	Acc2	Organism
9	39	3.9	2712	94	AF308696	AF308696	Mus muscu
10	35	3.5	34593	94	AF348157	AF348157	Mus muscu
11	23	2.3	105490	89	AF165142	AF165142	Homo sapi
12	23	2.3	190705	65	AC019331	AC019331	Homo sapi
13	21	2.1	64451	90	AL139399	AL139399	Human DNA
14	20	2.0	1176	8	AB046207	AB046207	Anguilla
15	20	2.0	1573	14	U23796	U23796	Arabidopsis
16	20	2.0	4918	12	AF047031	AF047031	Arabidops
17	20	2.0	6718	85	AB014516	AB014516	Homo sapi
18	20	2.0	46275	85	AC003107	AC003107	Human DNA
19	20	2.0	79590	12	AB007727	AB007727	Arabidops
20	20	2.0	84129	12	AB013392	AB013392	Arabidops
21	20	2.0	103638	13	ATF14F18	ATF14F18	Arabidops
22	20	2.0	117338	92	HS173D1	HS173D1	Human DNA
23	20	2.0	148508	71	AL031984	AL031984	Human DNA
24	20	2.0	149180	63	AC027795	AC027795	Homo sapi
25	20	2.0	150683	74	AC015525	AC015525	Homo sapi
26	20	2.0	165493	65	AC069420	AC069420	Homo sapi
27	20	2.0	168997	65	AC018876	AC018876	Homo sapi
28	20	2.0	177868	69	AC021927	AC021927	Homo sapi
29	20	2.0	183412	83	AP001899	AP001899	Homo sapi
30	20	2.0	183855	61	AC025191	AC025191	Homo sapi
31	20	2.0	184855	67	AC022446	AC022446	Homo sapi
32	20	2.0	192059	77	AC090095	AC090095	Homo sapi
33	20	2.0	195583	62	AC01804	AC01804	Homo sapi
34	20	2.0	197788	80	AL358492	AL358492	Homo sapi
35	20	2.0	199812	65	AC018443	AC018443	Homo sapi
36	20	2.0	222876	73	AC068379	AC068379	Homo sapi
37	20	2.0	222876	73	AC068379	AC068379	Homo sapi
38	20	2.0	235141	85	AC004615	AC004615	Homo sapi
39	19	1.9	2051	95	AF178954	AF178954	Mus muscu
40	19	1.9	2118	95	RNU87627	U87627	Rattus norv
41	19	1.9	5841	13	AF136600	AF136600	Magnaport
42	19	1.9	42545	6	CELC32B5	U80843	Caenorhabdi
43	19	1.9	43390	86	AC007191	AC007191	Homo sapi
44	19	1.9	47404	91	AP000333	AP000333	Homo sapi
45	19	1.9	61435	78	AC091020	AC091020	Homo sapi

```
LOCUS       AF304370     2908 bp    mRNA            PRI       23-FEB-2001
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 mRNA, complete cds.
ACCESSION   AF304370
VERSION     AF304370.1  GI:1080932
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED    11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     CDS             1..2481
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG244441.1"
                     /db_xref="GI:10880933"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERRKDPLRHLRTRE
                     KRGPSGCSGGPNTVIQVVAAGSRDSGAALIVFSEPNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKIYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPHSEQRGKHQPWQSPRPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVAFICKLHLKRGNFLVLKAKEMGLPVGTRAIAPIIA
                     AVKDGKSITHEGREILAEELCTPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMEERGPDTQHLVLNENCASVHNLRSHKICQTQLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPREWRQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAERRSQYPEITIFGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVAPNQLKAWLQQYRNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFCCALVHTSGWKVVYSGDTMPCEALVRGMKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRNAEFIMLNHFSQRYAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDPPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPOQKRAHTEEPOAKKVRAQ"
BASE COUNT      702 a    793 c    826 g    587 t
ORIGIN
```

ALIGNMENTS

RESULT 1

AF304370

```
Query Match  87.9%;   Score 880;   DB 89;   Length 2908;
Best Local Similarity  99.9%;   Pred. No. 0;
Matches 1000;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy     1  tcgcagccacaagagattcaaaacccagttcaaactccatccacccggacaatcttccccctgct  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1140  TCGCAGCCACAAGAGATTCAAAACCCAGTTCAAACTCCATCCACCCGGACAATCTTCCCCCTGCT  1199

Qy    61  caccagtttccgctgtaagagaagagaggcccaccctcagtgtgcccatggttcagggtga  120
          |||||||||| ||||| ||||| |||||||||| |||||||||||||||||||||||||| ||
Db  1200  CACCAGTTTCCCCTGTRAAGAAGGAGGGCCCCAACCTCAGTGTGCCCATGGTTCAGGGTGA  1259

Qy   121  atgcctcctcaagtaccagctccgtcccaggagggagggagggagggatgccattattac  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1260  ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGGGAGGGAGGGATGCCATTATTAC  1319

Qy   181  ttgcaatccctcaggaattcatcatagttcaggcgctccagagggcgtgcagaggcgt  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1320  TTGCAATCCCTCAGGAATTCATCATAGTTCAGGCGCTCCAGAGGGCGTGCAGAGACGT  1379

Qy   241  gcaggagtacaggagagagtgcgcaggagacggcgcccaagcccagcagagaaagaagtcagta  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1380  GCAGGAGTACAGGAGAGAGTGCGCCAGGAGACGGCCCCAGCCCAGCAGAGAAAAGAAGTCAGTA  1439

Qy   301  cccagaaatcatcttccttggaacagggtctgccatccgatgaaagattcgaaatgtcag  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1440  CCCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCGATGAAAGATTCGAAATGTCAG  1499

Qy   361  tgccacacttgtcaacataagcccgacacgtctctgctactggactgtgtgagggcac  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1500  TGCCACACTTGTCAACATAAGCCCGACACGTCTCTGCTACTGGACTGTGTGGGGGCAC  1559

Qy   421  atttgggcagctgtgccgtcattacggagaccagggtcctgggcaccctggc  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1560  ATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGGTGGACAGGGTGCTGGGCACCCTGGC  1619
```

```
Qy    481  tgctgtgttgttgtccaacctggcacggacggcagatcaccacacgggcttgccaagtatcttgc  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620  TGCTGTGTTTGTTGTCCACCT-GCACCGAGATCACCACACGGGCTTGCCAAGTATCTTGC  1678

Qy    541  tgcaggagaacgcgccttggccatcttcttgggaaagccgcttcacccttgctggtggttg  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1679  TGCAGAGAGAACGCGCCGTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTG  1738

Qy    601  ccccaaccacgctcaaagccctggctccagcagtaccacaacagtgccagaggctcctgc  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1739  CCCCCAACCAGCTCAAAAGCCCTGGCTCCAGCAGTACCACAACAGTGCCAGAGGTCCTGC  1798

Qy    661  accacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctg  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1799  ACCACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTG  1858

Qy    721  cagtggaaagattgatcagttcgctgttgtgcgaacatgtgattcggaaagttcagacct  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1859  CAGTGGAAAAGATTGATCAGTTCGCTGTTGTGCGAACACATGTGATTCGGAAAGTTCAGACCT  1918

Qy    781  gtctggtgtgcgcacctccaagcaatgtcgttttgtctgtcgctgtgtgcacacctctggca  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1919  GTCTGGTGTGCGGCACTGCAAGCAATGCGTTTGTCTGTCGCTGTGTGCCACACCTCTGGCTGA  1978

Qy    841  aagtggtctattccgggacaccacatgccctgccgaggctctggtccggatgggaaagtg  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1979  AAGTGGTCTATTCCGGGACACCATGCCCTGCCGAGGCTCTGGTCCGGATGGGAAAGATG  2038

Qy    901  ccacccttcctgatcatgaagccaccctgggaagatgtgtttggaagagaagcagtggaa  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2039  CCACCCTCCTGATCACATGAAGCCACCCTGGGAAGATGGTTTGGAAGAGAAGCAGTGGAAA  2098

Qy    961  agacacacagacaaagtcccaagccaatcacgcgtgtgggatg  1001
           ||||||||||||||||||||||||||||||||||||||||||
Db   2099  AGACACACAGCACAACGTCCCAAGCCAATCAGCCGTGGGGATG  2139
```

```
RESULT   2
LOCUS       BC001939                2997 bp    mRNA              PRI      16-MAR-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:4102, mRNA, complete cds.
ACCESSION   BC001939
VERSION     BC001939.1    GI:12804972
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2997)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
```

```
            Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: i Column: 17.
FEATURES             Location/Qualifiers
     source          1..2997
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:4102 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             61..2541
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH01939.1"
                     /db_xref="GI:12804973"
                     /translation="MWALCSLIRSAAGRIMSQGRTISQAPARERRPRKDPLRIHLRTRE
                     KRPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWHWSNVGGLSGMIITLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGHQPWQSPERLSRLSPERSDSELN
                     ENEPHLPHGVSQRGVRDSSLVVAFICKLHLKRGNFLVKAKEMGLPVGTAAIAPIA
                     AVKDGKSITHEGRILAEELCTPDPGAAFVVVECPDESFIQPICENAFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFQPDTOHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIFPLITSFRCKKEGPFLSVPNVQGECLKLYQLRPREEWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQQPAPAEKRSQYPEIIFLGTSSAIPMKIRNVSATLVNISP
                     DTSLLLDGGEGTTGQLCRHYGDQVDRVLGTLAAVFEVSHLHADHHITGLPSILLORERAL
                     ISSLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHIISMIPAKCLQEGAIISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEAAVEKTHSTTSOAISVGMRNNAEFIMLNHSQRYAKVPLSPNFS
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMERREKRELRQVRAALLSRE
                     LAGLEDGEPQOKRAHTEEPQAKKVRAQ"
BASE COUNT      806 a    853 g    598 t
ORIGIN
            740 a    806 c    853 g    598 t
```

```
Query Match      82.8%;  Score 829;  DB 91;  Length 2997;
Best Local Similarity   99.8%;
Matches  999;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy      1  tcgcagccacaaagattcaaacccagctcaacctcatccaccggagacatctccccctgct  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200  TCGCAGCCACAAAGATTCAAAACCCAGCTCAACCTCATCCACCACCGGACATCTCCCCCTGCT  1259

Qy     61  caccagttccgctgtagaaagagaggcccaccaccctcagtgtgccatggttcagggtga  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260  CACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCATGGTTCAGGGTGA  1319

Qy    121  atgcctcctcaagtaccagagctccgtcccaggagggaggggagtggcagggatgccattactac  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320  ATGCCTCCTCAAGTACCACAGCTCCGTCCCAGGAGGAGGTGGCAGAGGGATGCCATTATTAC  1379

Qy    181  ttgcaatcctgaggaagaattcatagttgaggcgctgcccagcttcccaacttccagcagcgt  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380  TTGCAATCCTGAGGAGAATTCATAGTTGAGGCGCTGCCCAGCTTCCCAACTTCCAGCAGCGT  1439

Qy    241  gcaggagtacaggaggagtgcgcagagacgcgccagccacccagccagaagaaagaagtcagta  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440  GCAGGAGTACAGGAGGAGTGCGCAGAGACGCGCCACCCCAGCCAGAAGAAAGAAGTCAGTA  1499

Qy    301  cccggaaatcatcttccttggaacagggtctgcatccggatgaagattcgaaatgtcag  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500  CCCAGAAATCATCTTCCTTGGAACAGGGTCTGCATCCGGATGAAGATTCGAAATGTCAG  1559

Qy    361  tgtccacacttgtcaacataagcccgacacagtatctgctactggactgtggtggggcac  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560  TGCCACACTTGTCAACATAAGCCCGACACAGTCTCTGCTACTGGACTGTGGTGGGGCAC  1619

Qy    421  attggggcagctgtgccgtcattacggagacaggtggacagggtcctggggcaccctggc  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT   3
BC004158

```
Qy   1620  GTTTGGGCAGCTGTGCCGTCATTACGGACGACCAGGTGGACAGGGTGCCTGGGCACCCTGGC  1679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680  TGCTGTGTTTGTGTCCCACCT-GCACGCAGATCACCACACGGGGCTTGCCAAGTATCTTGC  1738

Qy    481  tgctgtgtttgtgtcccacctggcacgcagatcaccacacggggcttgccaagtatcttgc   540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680  TGCTGTGTTTGTGTCCCACCT-GCACGCAGATCACCACACGGGGCTTGCCAAGTATCTTGC  1738

Qy    541  tgcaggagaacgcgcctggcatcttggaaagccgcttcacccttgctggttggttg        600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1739  TGCAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTTG       1798

Qy    601  cccccaaccaagctcaaagctggctccagcagtaccacacaccagtgccaggagg tcctgc   660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1799  CCCCCAACCAACTCAAAGCCTGGGCTCCAGCAGTACCACACACCAGTGCCAGGAGGTCCTGC  1858

Qy    661  accacatcagtatgattcctgccaaaatgcctt caggaagaagggggcttaagatctccactg   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1859  ACCACATCAGTATGATTCCTCGCCAAAATGCCTTCAGGAAGAGGGGCTGAGATCTCCACTG   1918

Qy    721  cagtggaaaagattgatcagttctcgctgttgcgaacatgtgattggagagtttcagacct   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1919  CAGTGGAAAAGATTGATCAGTTCTCGCTGTTGCGAACATGTGATTGGAGAGTTTCAGACCT   1978

Qy    781  gtctgtgtgtcggcactgcaagacgtcgtttggctgtcgcagtgtcacacctctgctgga   840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1979  GTCTGTGTGCGGCACTCCAAGCATGCGTTTGGCTGTCGCGTCGGTGCACACCTCTGGCTGGA  2038

Qy    841  aagtggtctattccgggggacacatgcccctgcgcaggcgctcggtcggagtggggaaagtg   900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2039  AAGTGGTCTATTCCGGGGACACATGCCCTCGCTCGGAGCGCTCCTGCTGTGCGGGAAAGATG  2098

Qy    901  ccacctcctgataacatgaagccacccctgggaagatggtttggaagagaggaggagcagtcgaaa   960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2099  CCACCCTCCTGATACATGAAGCCACCCCTGAGAAGATGGTTTGGAAGAGAGGAGGAGCAGTCGGAAA  2158

Qy    961  agaacaacagacaacgtcccaagccatcagcgtgggagtg 1001
           |||||||||||||||||||||||||||||||||||||||
Db   2159  AGAACAACAGCAACGTCCCAAGCCATCACGGTGGGGATG 2199
```

FEATURES   Location/Qualifiers
     source   1..3006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:2441 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS   70..2550
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH04158.1"
                     /db_xref="GI:13278771"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSCCSGGPENTVLLQVVAAGSRDSGAALIVFSENRYLENCDGEVQVRLMQEHKLKV
                     ARLDNIFLTRMHIWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVIQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRRGVRDSSLVVAAFICKLHLHKRGNFLVLKAKEMGLPVGTRAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVWECDDESFIQPICENATFQRYOGKRAD
                     APVALVVHMAPASVLVDSRYQQWMEREGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTISFRCKKEGPTLSVPMVGGECLLKYQLRPRREWQRDALIITCNPEEFIVEA
                     LQLPNFQQSVGEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNTSP
                     DISLLLDCGEGTFGQLCRHYGDVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKANLQQYNLHIHSMIPAKCLCGEAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFCCALVHTSGWRKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDFPTMRKLIPLKALFAGDIEEMEREREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"

BASE COUNT   742 a   808 c   858 g   598 t
ORIGIN

```
Query Match      82.8%;    Score 829;    DB 91;    Length 3006;
Best Local Similarity    99.8%;
Matches 999;   Conservative    0;    Pred. No. 0;    Indels    1;    Gaps    1;
                Mismatches    1;

Qy     1  tcgcagccacaagattcaaacccagctcaacctcatccaccoggacaatcttcccctgct   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1209  TCGCAGCCACAAGAGTTCAAAACCCAGCTCAACCTCATCCACCCGGACAATCTTCCCCTGCT  1268

Qy    61  caccagttccgctgtaagaaggaggggccccaccccaagctcaagtgtcccaatggtgga  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1269  CACCAGTTTCCCCTGTAAGAAGGAGGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGA  1328

Qy   121  atgcccctcaagtacagcctccgtccccagccgagggggatggcaggagggaatggcattac  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1329  ATGCCCTCCTCCAAGTACCAGCTCCGTCCCCAGGACGGAGGAGTGGCACAGAGGGCCATTATTAC  1388

Qy   181  ttgcaatcctgagggaattcatagttgagcgcctgcagcttcccaacttccagcagagcgt  240.
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1389  TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT  1448

Qy   241  gcaggagtacagaggaggaggtgcgcaggagaggcgacccaggcagcaggagagcagta  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1449  GCAGGAGTACAGAGGAGGTGCCAGGACGGCCCAGGCACCCCAGCCCCAGGAGAGAAAGAAGTCAGTA  1508

Qy   301  cccagaaaatcatcttccttggaacaggggttctgccacacccggtgaaggattgaaaatgtcag  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1509  CCCAGAAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGAATCGAAATGTCAG  1568

Qy   361  tgccacacttgtcaacataagcccccgacacgtctctgctactggactgtggtgagggcac  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1569  TGCCACACTTGTCAACATAAGCCCCCGACACACGTCTCTGCTACTGGACTGTGGTGAGGGCAC  1628

Qy   421  atttgggcaagctgtgcgtcattacggagaccaagttcctgggaacctgggc  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1629  GTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGC  1688

Qy   481  tgctgtgtttgtgtcccaccggcagatcaccacacggggcttgccaagtatcttgc  540
```

```
RESULT  4
LOCUS       AK001392    2976 bp    mRNA            PRI        22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION   AK001392
VERSION     AK001392.1    GI:7022621
KEYWORDS    .
SOURCE      oligo capping; fis (full insert sequence).
            Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2000985.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2976)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812 Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
  COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /clone="NT2RP2000985"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP2"
                     /note="Cloning vector: pME18SFL3; mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             75..2639
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91666.1"
                     /db_xref="GI:7022622"
                     /translation="MWALCSLLRSAAGRTMSGGRTISQAPARERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALIYPSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHNHSNVGGLSGMLILTKETGLPRCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAMRPHISAPEYEDEMTVVIQIPIIHSEQRGKHQPWQSPERPLSRLSPERSDSDSSEN
                     ENEPHLPHGVSQRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAIAPITA
                     AVKDGKSITHEGREILAFELCTPPDPGAAFVVVECPDESFIQPICENATFQRYGKAD
                     APVALVVHMAPASVLVDSRYQOWMERGPDTOHLVLNENCASVHNLRSHKIQTQLNLI
                     HDIFPLLTSFRCKEGPPLLSVPWVQGECLLKTQVTLGCASAIPMKIRNVSATLVNISPD
                     QLPNFQQSVQEYRRSAQQGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISPD
                     TSLLLDGGEGTGSQLCHIYGDQVDRVLGTLAAFVSHLHADHHTGLPSILLORERALA
                     SLGKPLHPLLVAPNQLKAWLQQYYNQEOVLHHISMIPAKCLQEGAEISSPAVERLI
                     SSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATLL
                     IHEATLEDGLEEEAVEVKTHSSTISQAISVGMRNNAEFIMLNHSQRYAKVPLFSPNFSE
                     KVGVAFFDIMKVCLGDFFPMPKLIPPLKALFAGDIEEMEERREKRELQVRAALLSREL
                     AGLLEDGSLSRSGPTQRSHRPRRSEPSEDLGPDELRRLCVFCPTHAPVSALLAGRS"
BASE COUNT      712 a    807 c    856 g    601 t
ORIGIN
```

```
Qy      80 aaggagggccccacccctcagttgtgcccatggttcaaggtgaatgcctctcaagtaccag 139
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1290 aaggagggccccacccctcagttgtgcccatggttcaaggtgaatgcctctcaagtaccag 1349

Qy     140 ctccgtcccaggagggagtggcagagggatgcacattattacttgcaatcctgaggaattc 199
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1350 ctccgtcccaggagggagtggcagagggatgcacattattacttgcaatcctgaggaattc 1409

Qy     200 atagttgaggcgctgcagcttcccaacttccaagcacagagcgtgcaggagtacaggaggagt 259
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1410 atagttgaggcgctgcagcttcccaacttccaagcacagagcgtgcaggagtacaggaggagt 1469

Qy     260 gcgcaggacgacgcccagcccagcagagaaagaaagtcagtaccaggaaatcatcttcctt 319
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1470 gcgcaggacgacgcccagcccagcagagaaagaaagtcagtaccaggaaatcatcttcctt 1529

Qy     320 ggaacaggtgctgccatcccgatgaagattcgaaatgtcagtgccacacttgtcaaacata 379
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1530 ggaacaggtgctgccatcccgatgaagattcgaaatgtcagtgccacacttgtcaaacata 1589

Qy     380 agccccgacacgtctctcgctactggagatgtggaggggcacatttggagcagtgtgccgt 439
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1590 agccccgacacgtctctcgctactggagatgtggaggggcacatttggagcagtgtgccgt 1649

Qy     440 cattacggagaccaaggtggacaggtcctgggcacccctggcctgtctgtgttgtcccac 499
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1650 cattacggagaccaaggtggacaggtcctgggcacccctggcctgtctgtgttgtcccac 1709

Qy     500 ctggcacgcagatcaacacacgggcttgccaagtatcttgctgcagagaacggcgcctt 559
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1710 ct-gcaccgcagatcaacacacgggcttgccaagtatcttgctgcagagaacggcgcctt 1768

Qy     560 ggcatcttgggagaagcgcttcacccctttgctggtgttgcccccaaccagtcaaagc 619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1769 ggcatcttgggagaagcgcttcacccctttgctgtgttgtgcccccaaccagtcaaagc 1828
```

```
Qy     961 agacacacacagcaaagtcccaagccatcagcgtggggatg 1001
           ||||||||||||||||||||||||||||||||||||||||
Db    2168 agacacacagcgacacgtccaagccatcacggtgggggatg 2208

Qy     901 ccaccctcctgataacatgaggcaccacccctgaagaataggtgttggaaggagaagcagtggaaa 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2108 ccaccctcctgatacatgaagcccaccctgaagaataggtgtttggaagaggaagcagtggaaa 2167

Qy     841 aagtggtctattccggggacaccatgccctggcgaggctctgtccggatgggggaagagatg 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2048 aagtggtctattccggggacaccatgccctccgaggctctgtccggatgggggaaagagatg 2107

Qy     781 gtcttggtgtgcggcactgcaagacatgtcgttgcgctggtgcgcaccctctgctgaa 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1988 gtctgtgtgcgcactgcaagacatgtcgttgcgctggtgcgcaccctctgctgga 2047

Qy     721 cagtggaaaaagattgatcagtctcgtgttgcgaacatgtgatttgaagagttcagacct 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1928 cagtggaaaagattgatcagtctcgtgttgcgaacatgtgatttgaagagttcagacct 1987

Qy     661 accacatcagtatgatgattcctggcaaatgccttcaggagggggctgagaatccagtcctg 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1868 accacatcagtatgatgattcctggcaaatgccttcaggagggggctgagaatccagtcctg 1927

Qy     601 ccccaaccagctcaaagctctggctccagcagtaccacacagtgccaggaggtcctgc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1808 ccccaaccacctcaaagctctggctccagcagtaccacacagtgccaggaggtcctgc 1867

Qy     541 tgcaagagggaacgcgccttggcatcttgggaaagccgcgtcaccctttgttgtggttg 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1748 tgcagagagaacgcgcccttggcatcttgggaaagccgcgtcaccctttgttgtggttg 1807

Qy    1689 tgctgtgttgtttgtccccacct-gcaccgagatcaccaccacgggctttgccaagtatcttgc 1747
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1748 TGCAGAGAGAACGCGCCCTTGGCATCTTTGGGAAAGCCGCGTCACCCTTTGTTGTGGTTG 1807
```

```
Qy    620  ctggtccagtaccacaaccagtgccaggagtcctgcacacacacactcagtatgatcc  679
           ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db   1829  CTGGCTCCAGCAGTACTACAACCACAGTGCCAGGAGGTCCTCCACACACACAGTTCC  1888

Qy    680  tgccaaatgccttcaggaagggggctgagatctccagtcctgcagtggaaagattgatcag  739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1889  TGCCAAAATGCCTTCAGGAAGGGGGCTGAGATCTCCAGTCCTGCAGTGGAAAATTGATCAG  1948

Qy    740  ttccgtgttgcgaacatgtgatttggaaagagtttcagacctgtctggtgcggcactgcaa  799
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1949  TTCCGTGTTGCGGAACATGTGATTTGGAAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAA  2008

Qy    800  gcatcgtcgtttggctgtgtgcgctggtgcacacctctgggaaagtggttcattccgggga  859
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2009  GCATCGTCGTTTGGGCTGTGTGCCCTGGTGCACACCTCTCGGCTGGAAAGTGGTTCATTCCGGGGA  2068

Qy    860  caccatgtgccctcgcgaggctctggtccggatctggaggaaagtgccaccctcctgatacatga  919
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2069  CACCATGCCCTGCCGGAGGCTCTGGTCCGGATGGGAAAGAGTGCCACCCTCCTGATACATGA  2128

Qy    920  agcacaccctggaaagatggtttgggaaggaggaagcagtgggaagaacacacaacaaagtc  979
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2129  AGCCACCCTGGAAGATGGTTTGGAAGAGGAGCAGTGGAAAAGAGACACACAACAACGTC  2188

Qy    980  ccaagcccatcaagcgtggggatg  1001
           |||||||||||||||||||||||
Db   2189  CCAAGCCCATCAGCCGTGGGGATG  2210
```

RESULT 5

```
LOCUS        AF308698               2908 bp    mRNA             PRI       27-FEB-2001
DEFINITION   Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION    AF308698
VERSION      AF308698.1  GI:10464496
KEYWORDS     .
SOURCE       chimpanzee.
  ORGANISM   Pan troglodytes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE    1  (bases 1 to 2908)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
             Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
             Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
             Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
             Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE      A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL    Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED     11175785
REFERENCE    2  (bases 1 to 2908)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
             Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
             Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
             Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE      Pan troglodytes ortholog of human HPC2/ELAC2
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 2908)
  AUTHORS    Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
             Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
             Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
             Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
             Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
             Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
             Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Swensen,J.,
```

```
  TITLE      Direct Submission
  JOURNAL    Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
             Lake City, UT 84103, USA
```

Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.

```
FEATURES             Location/Qualifiers
     source          1..2908
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     CDS             1..2481
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24920.1"
                     /db_xref="GI:10464497"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCCSGGPNTVLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGIQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITIEGPEILLAEELCTPPDGEQRVAEVVVECPGEDSFIQPICENAAFPYQGKAD
                     APVALVVHHMAPESVLVDSRYQQWMRFGPDIQHLVLNENCASVHNLRSHKIQTOLNLI
                     HPDIFPLLTSFPCKKEGPTILSVPMVOGECLLKYQILRPREMQRDAIITCNPEEFIIEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLEFGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGESTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLNILLQRERAL
                     ASLGKPFHPLLVAPNQLKAWLQQYHNCQQEVLLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHICKHAFGCALYHTSGWKVVYSGDTMPCEALVRMGRDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSORYAKVPLFSPNEN
                     EKVGVAFDHMKVCFGDFATMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEEPQAAKVRAQ"
BASE COUNT      712 a    788 c    819 g    589 t
ORIGIN
```

```
Query Match           37.18;   Score 371;   DB 89;   Length 2908;
Best Local Similarity  99.8%;   Pred. No. 2.9e-205;
Matches  421;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;
```

```
Qy    580  ttcaacccttcgctggttggttgccccccaacaagctcaaaagcctggcctcacggaag  639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1718  TTCACCCTTGCTGGTTGGTTGCCCCCCAACAGCTCAAAAGCCTGGCCTCCAGCAGTACCACA  1777

Qy    640  accagtgccacaggagtcctgcaccacatcagtatgatcctgccaaaatgcctcaggaag  699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1778  ACCAGTGCCACAGGAGGTCCTGCACCACATCAGTATGATCCTGCCAAAATGCCTTCAGGAAG  1837

Qy    700  gggctgagatctccagtcctgcagtggaaagagattgatcagttcgctgttgcgaacatgtg  759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1838  GGGCTGAGATCTCCAGTCCTGCAGTGGAAAGAGATTGATCAGTTCCGTGCTTGCGAACATGTG  1897

Qy    760  atttggaagagtttcagacctgtctggtgcggcactgcaagcaagcatgtctttgggctgtcgc  819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1898  ATTTGGAAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCAAGCCATGGCTGTTTGGCCTGTCGC  1957

Qy    820  tggtgcacacctctggctggaaagtggtctattccggggacaacatgccctgcgcgagctc  879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1958  TGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACACCATGCCCTGCCGAGCCTC  2017

Qy    880  tggtccggatgggaaagatgccaacctcctgatcaagcacatgggaagatggtt  939
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2018  TGGTCCGGATGGGGAAAGATGCCACCCTCCTGATCAAGCACACCCTGAAGAGACGGTT  2077

Qy    940  tggaaagaggaggcagtggaaagaacacacaacaacgtcccaagcccatcagcgtggggga  999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2078  TGGAAGAGGAGGCAGTGGAAAAGAGACACACAACAACGTCCCAAGCCCATCAGCGTGGGGA  2137

Qy   1000  tg  1001
           ||
Db   2138  TG  2139
```

RESULT 6

```
LOCUS        AF308694               2893 bp    mRNA             PRI       27-FEB-2001
DEFINITION   Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
```

```
ACCESSION   AF308694
VERSION     AF308694.1  GI:10946488
KEYWORDS    .
SOURCE      Gorilla gorilla
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Gorilla gorilla ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1..2893
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
     gene            1..2893
                     /gene="ELAC2"
     CDS             1..2481
                     /gene="ELAC2"
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24916.1"
                     /db_xref="GI:10946489"
                     /translation="MWALCSLLRSAAGRTMSGGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCGGGNTVYLQVVAAGSRDSGAALYVFSENRYLNCGEGVQRLMQEHKLKV
                     VRLDNIFLTRMHWSNVGGLSGMILLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDTMTVYQIPIHSEQRGRHQPWQSPRPLSRLSPERSSDSSESN
                     ENEPHLPHGVSQRRGYRDSSLVVAFICKLHLKRGNEIVLKAKEMGLPVGTANAPITA
                     AVKDGKSITHEGREILAEELCTPPDGAAFVVVECPDESFIQPICENATFQRYQGRAD
                     APVALVWHMAPESVLVDSRYQOWMERGPDTDVRLGTLNENCASVHNLRSHKIQTQLNI
                     HPDIFPLLTSPPCKKEGPTLSVPMVQGECLLKYQLRPREWORDAIITCNPEEFIVEA
                     LQLPNPEQOSVQOEYRSVQOVDVPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSAPLVNISP
                     DTSLLDCGEGTFGQLCRHYGDVDRVLGTLAAVEFSHLHADHHTGLLNILLQREQAL
                     ASLGKPLHPLLVVAPSQLKAWLQQYINQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTVHNSFHSQAISVGMHMNAEFIMLNHSQRYAKVPLFSPNFN
                     EKVGVAPDHMKVCFGDPPTMPKLIPPLKALFAGDIEMEEREREKRELRQVRALLSGE
                     LAGGLEDGEPOQKRAHTEEPQAKKVRAQ"
```

```
BASE COUNT      705 a    787 c    815 g    586 t
ORIGIN

Query Match    34.3%; Score 343; DB 89; Length 2893; Best Local
Similarity 99.7%; Pred. No. 7.2e-189;
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy    608 ccagctcaaagcctggctccagcagtaccacaccacaagtgccaggaggtcctgcaccat 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1746 CCAGCTCAAAGCCTGGCTCCAGCAGTACCACACCACAAGTGCCAGGAGGTCCTGCACCAT 1805

Qy    668 cagtatgattcctgccaaatgccttcaggaaggggctgaagatctccaatcctgcagtgga 727
          |||||||| |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   1806 CAGTATGGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAAGATCTCCAGTCCTGCAGTGGA 1865

Qy    728 aagattgatcaagttcgctgttcggaacatgtgattggaaagagttccgaacctgtctggt 787
          ||||||||||||| ||||||||||||||||||||||| ||||||| ||||||||||||||||
Db   1866 AAGATTGATCAGTTCGCTGTTCGGAACATGTGATTGGAAGAGTTTCAGAACCTGTCTGGT 1925

Qy    788 gcggcacctcgaaagcatgcggtttggctgtgcgtggtgcacccctctggttggaaagtggt 847
          |||||||| |||||||||||||||||| |||||||||||||| ||| ||||| |||||||||
Db   1926 GCGGCACTCCAAAGCATGCGGTTTGGCCTGTGCGTGGTGCACACCTTCTGGCTGGAAAGTGGT 1985

Qy    848 ctattccgggacacatcgccctgcgcgaggctctcggtcggtggggaaagatgccaccct 907
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1986 CTATTCCGGGACACCATCGCCCTGCGCGAGGCTCTCGGTCGGTGGGGAAAGATGCCACCCT 2045

Qy    908 cctgatacatgaagccacccctggaagatggtttggaagaggaagaagacaca 967
          ||||||||||||||||||||||||||||||||||||| ||||| || ||||||
Db   2046 CCTGATACATGAAGCCACCCCTGGAAGATGGTTTGGAGAGGAAGGAAGACAGTGGAAAAAGACACA 2105

Qy    968 cagcacaaagtcccaagccatcagcgtggggatg 1001
          ||||||||||||||||||||||||||||||||||
Db   2106 CAGCACAAAGTCCCAAGCCATCAGCGTGGGGATG 2139
```

```
RESULT 7
ACCESSION   AC005277
LOCUS       AC005277    118788 bp    DNA    PRI    23-JUL-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
ACCESSION   AC005277
VERSION     AC005277.1  GI:3337311
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.597_M_12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
```

```
          Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Bortin,C.,
          Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
          Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
          FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
          Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
          Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
          Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
          Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
          Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
          Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
          Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
          Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
          Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
          Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jul 23, 1998 this sequence version replaced gi:3335015.
          All repeats were identified using RepeatMasker: Smit,A.F.A. &
          Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html

          Only the first 118.8 kilobases of this clone are being submitted.
          The remainder overlaps accession number AC005274 (WICGR project
          L350).
FEATURES             Location/Qualifiers
     source          1..118788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="hRPK.597_M_12"
                     /map="17"
                     /clone_lib="RPCI-11 human BAC library"
     repeat_region   795..884
                     /rpt_family="MIR"
     repeat_region   complement(932..1000)
                     /rpt_family="MIR"
     repeat_region   2044..2076
                     /rpt_family="AT_rich"
     repeat_region   2237..2713
                     /rpt_family="AT_rich"
     repeat_region   2718..3477
                     /rpt_family="MER34"
     repeat_region   complement(3478..3821)
                     /rpt_family="L2"
     repeat_region   3822..4007
                     /rpt_family="MLT1A1"
     repeat_region   4015..4281
                     /rpt_family="L2"
     repeat_region   complement(4282..4321)
                     /rpt_family="L2"
     repeat_region   complement(4328..4608)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(4609..4895)
                     /rpt_family="AluSx"
     repeat_region   complement(4896..5186)
                     /rpt_family="MER89"
     repeat_region   5219..5479
                     /rpt_family="AluSp"
     repeat_region   complement(5481..5740)
                     /rpt_family="L2"
     repeat_region   complement(5963..6275)
                     /rpt_family="MER34"
     repeat_region   complement(6411..6494)
                     /rpt_family="AluSp"
     repeat_region   6830..7013
                     /rpt_family="MER5A"
     repeat_region   7014..7318
                     /rpt_family="MLT1D"
     repeat_region   7319..7581
                     /rpt_family="AluYb8"
     repeat_region   7721..8032
                     /rpt_family="MLT1C"
     repeat_region   8039..8333
                     /rpt_family="AluSq"
     repeat_region   9431..9460
                     /rpt_family="MSTB"
     repeat_region   9662..9829
                     /rpt_family="AT_rich"
     repeat_region   complement(11302..11728)
                     /rpt_family="MIR"
     repeat_region   complement(11824..11884)
                     /rpt_family="MER39"
     repeat_region   complement(11928..12108)
                     /rpt_family="LTR16C"
     repeat_region   12453..12725
                     /rpt_family="AluJo"
     repeat_region   13377..13785
                     /rpt_family="AluJb"
     repeat_region   14118..14153
                     /rpt_family="L2"
     repeat_region   complement(15320..15620)
                     /rpt_family="AluJo"
     repeat_region   15683..15759
                     /rpt_family="(CA)n"
     repeat_region   complement(15838..15893)
                     /rpt_family="MER91A"
     repeat_region   complement(15915..16080)
                     /rpt_family="MER91A"
     repeat_region   complement(16474..16544)
                     /rpt_family="MER91A"
     repeat_region   17907..17920
                     /rpt_family="L2"
     repeat_region   18271..18311
                     /rpt_family="L1ME"
     repeat_region   complement(18368..18529)
                     /rpt_family="L2"
     repeat_region   complement(18530..18560)
                     /rpt_family="AluJb"
     repeat_region   complement(18561..18681)
                     /rpt_family="(CA)n"
     repeat_region   18760..19065
                     /rpt_family="AluJb"
     repeat_region   complement(19638..19896)
                     /rpt_family="AluSg"
     repeat_region   complement(19971..20260)
                     /rpt_family="MIR"
     repeat_region   20263..20552
                     /rpt_family="L1ME3A"
     repeat_region   complement(20553..20816)
                     /rpt_family="MER58B"
     repeat_region   complement(20821..21069)
                     /rpt_family="L1ME3A"
     repeat_region   complement(21071..21100)
                     /rpt_family="L1M4"
     repeat_region   complement(21101..21338)
                     /rpt_family="L1M4"
     repeat_region   complement(21348..21598)
                     /rpt_family="(TAAA)n"
     repeat_region   21799..22104
                     /rpt_family="AluSx"
     repeat_region   22151..22562
                     /rpt_family="L1M4"
     repeat_region   23978..24287
                     /rpt_family="AluJo"
     repeat_region   24852..25199
                     /rpt_family="MSTA"
     repeat_region   complement(25697..25985)
                     /rpt_family="AluSq"
     repeat_region   25816..25895
                     /rpt_family="L2"
     unsure          complement(25994..26099)
                     /note="Single-stranded coverage."
     repeat_region   complement(26122..26291)
                     /rpt_family="purine-rich"
     repeat_region   26312..26637
                     /rpt_family="MLT1D"
```

```
repeat_region   26456..26480
                /rpt_family="(CAAAA)n"
repeat_region
                /rpt_family="(GGGAA)n"
repeat_region   complement(26576..26781)
repeat_region   complement(26912..26990)
repeat_region   26991
                /rpt_family="MER6B"
repeat_region   26991..27290
                /rpt_family="MER21B"
repeat_region
                /rpt_family="AluSx"
repeat_region   complement(27291..27975)
repeat_region   28034..28197
                /rpt_family="MER21B"
repeat_region   28197
                /rpt_family="LTR28"
repeat_region   28266..28795
                /rpt_family="LTR28"
repeat_region   28843..28916
                /rpt_family="LTR28"
repeat_region   28920..29005
                /rpt_family="MIR"
repeat_region   29140..29233
                /rpt_family="MIR"
```

Query Match   12.4%; Score 124; DB 85; Length 118788;
Best Local Similarity 100.0%; Pred. No. 1.1e-60;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    769 agtttcagacctgctggtcggccgactgccaagacaatgtcgtttggctgtcgcgctggtgcaca 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101623 AGTTTCAGACCTGCTGGTCGGCCGACTGCCAAGACAATGTCGTTTGGCTGTCGCGCTGGTGCACA 101682

Qy    829 ccctggcctgggaaagtggtcttattccggggaccacatcgcccctggcgaggtctcggtccgga 888
          |||||||||||||||| ||||| |||||||||||||||||||||||||||||||| |||||||||
Db 101683 CCTCTGGCCTGGGAAAGTGGTCTATTCCGGGGACCACCATCGCCCCTGGCGAGGCTCTGGTCCGGA 101742

Qy    889 tggg 892
          ||||
Db 101743 TGGG 101746
```

RESULT 8
```
LOCUS       AF304371S2   740 bp   DNA   PRI   19-OCT-2000
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 17 and partial cds.
ACCESSION   AF304371S2
VERSION     AF304369.1  GI:10880929
KEYWORDS
SEGMENT     2 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Swensen,J., Schroeder,M.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.,
            and Cannon-Albright,L.A.
  TITLE     A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Swensen,J., Schroeder,M.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.,
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1..740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     mRNA            301..439
                     /number=17
     exon            301..439
                     /number=17
                     HPC2/ELAC2"
     CDS             <301..>439
                     /codon_start=2
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24440.1"
                     /db_xref="GI:10880931"
                     /translation="PDTSLLLDCCEGTFGQLCRHYGDQYDRVLGTLAAFVSHLHADH
                     HT"
BASE COUNT      154 a   188 c   173 g   225 t
ORIGIN
```

Query Match   12.3%; Score 123; DB 89; Length 740;
Best Local Similarity 100.0%; Pred. No. 4.1e-60;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    380 agcccgacacgtctctgtcactggactggtgaggtgaaggcacaattgggcagctgtgccgt 439
          |||||||||||||||||||||||||||||| |||||||| |||||||||||||||||||||||
Db    299 AGCCCCGACACGTCTCTGTCACTGGACTGTGGGTGAGGGCACAATTTGGGCAGCTGTGCCGT 358

Qy    440 cattacggaggaccaggtggacaggtcctggggcaccctggctgctgttgttgtgtcccac 499
          ||||||||||||||||||||||||| |||||| ||||||||||||||| ||||| |||||||
Db    359 CATTACGGAGGACCAGGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTGTTTGTGTCCCAC 418

Qy    500 ctg 502
          |||
Db    419 CTG 421
```

RESULT 9
```
LOCUS       AF308696   2712 bp   mRNA   ROD   27-FEB-2001
DEFINITION  Mus musculus ELAC2 mRNA, complete cds.
ACCESSION   AF308696
VERSION     AF308696.2  GI:11992378
KEYWORDS
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Snyder,S.C.,
            Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
```

RESULT 10
LOCUS   AF348157   34593 bp   DNA   ROD   04-APR-2001
DEFINITION Mus musculus putative prostate cancer susceptibility protein
            (Elac2) gene, complete cds, alternatively spliced.
ACCESSION   AF348157
VERSION   AF348157.1   GI:13540341
KEYWORDS   .
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Illev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE   Mouse Elac2-containing genomic DNA
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 34593)
AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Illev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE   Direct Submission
JOURNAL   Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES   Location/Qualifiers
     source   1..34593
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     mRNA   join(8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27801,27930..28039,28553..28652,
                     28733..28853,29101..29179,29710..>30639)
                     /gene="Elac2"
                     /note="alternatively spliced 24 exon form"
                     /product="putative prostate cancer susceptibility protein"
     mRNA   join(8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27801,27930..28039,28553..28652,
                     28733..28853,29101..29179,29710..>31277)
                     /gene="Elac2"
                     /note="alternatively spliced 25 exon form"
                     /product="putative prostate cancer susceptibility protein"
     gene   <8470..>31277
                     /gene="Elac2"
     CDS   join(8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16746..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27801,27930..28039,28553..28652,
                     28733..28853,29101..29179,29710..29710)
                     /gene="Elac2"
                     /note="24 exon form"
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"

REFERENCE   3 (bases 1 to 2712)
AUTHORS   Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
TITLE   Mouse ortholog of human HPC2/ELAC2
JOURNAL   Unpublished
REFERENCE   4 (bases 1 to 2712)
AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Illev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
TITLE   Direct Submission
JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
REMARK   Sequence update by submitter
COMMENT   On Dec 26, 2000 this sequence version replaced gi:10946492.
FEATURES   Location/Qualifiers
     source   1..2712
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="fetus"
     CDS   15..2510
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24918.2"
                     /db_xref="GI:11992379"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPSKDPLRHLRTREKRGPGP
                     GGPNFVYLQVVAAGGRDAGAALIYFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKIGIELAVRPH
                     SAPEYKDETWTVYQVPIHSERRCGKQOPSOSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRAWGRDPSLVAFAVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSIYEGREIAAEELCTFGPDGLVFIVVGDEGFIPICENDIFPKRYQAEADAPVA
                     LVHIAPESVLIDSRYQQWMERFGPDTQHLLINENCPSVINLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPVRGECLLKYQLRPKREWORDTPDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFCOLCRIYGQOIDRVLCSLTAVFSVSHLHADHHTGLLNTLLQRHALASIG
                     KPFQPLLVAPTOLRAWLQOYHNCQEILHHVSMIPAKCLQKGAEVSNTTLERLISL
                     LETCDLEEFQTCLVRHCKHAFGCALVHSSGMKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKHSHTTSOAIVNGMRMNAGFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPKRALFAGDIEEMVEWRERKRELRIVRAALLTQQADSP
                     EDREPOQKRAHTDEPHSPQSKKESVANTLGARV"

BASE COUNT   659 a   741 c   748 g   564 t
ORIGIN

Qy   887 gatgggaaagaatgccaccctcctgatactgaagccac 925
               |||||||||||||||||||||||||||||||||||||||
Db   2027 GATGGGGAAACAGATGCCACCCTCCTGATACATGAAGCCAC 2065

Query Match   3.9%; Score 39; DB 94; Length 2712;
Best Local Similarity   100.0%; Pred. No. 6e-11;
Matches   39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
RESULT 11
AF165142
LOCUS       AF165142     105490 bp    DNA             PRI       24-AUG-1999
DEFINITION  Homo sapiens chromosome 8 clone BAC 392C11 map 8p11-p12, complete
            sequence.
ACCESSION   AF165142
VERSION     AF165142.1  GI:5499748
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105490)
  AUTHORS   Schudy,A., Blechschmidt,K., Schillhabel,M., Baumgart,C., Menzel,U.,
            Weber,J., Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..105490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p11-p12"
                     /clone="BAC 392C11"
     repeat_region   333..592
                     /rpt_family="AluSx"
     repeat_region   716..761
                     /rpt_family="MIR"
     exon            complement(1299..1513)
                     /note="MZEF"
                     /evidence=not_experimental
     exon            3417..3519
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(3617..4219)
                     /rpt_family="Charlie1a"
     repeat_region   complement(4209..4252)
                     /rpt_family="Charlie1a"
     exon            4302..4405
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(5085..5193)
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(5315..5460)
                     /rpt_family="FLAM_C"
     repeat_region   complement(5347..5460)
                     /rpt_family="AluJo"
     exon            complement(5373..5537)
                     /note="MZEF"
                     /evidence=not_experimental
     exon            6488..6564
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   complement(6760..7029)
                     /rpt_family="AluY"
     repeat_region   7180..7227
                     /rpt_family="AluSx"
     repeat_region   complement(7232..7489)
                     /rpt_type=tandem
     repeat_region   complement(7489..7563)
                     /rpt_family="AluSx"
     exon            8092..8142
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            complement(8101..8149)
                     /note="Initial_exon"
     exon            8306..8360
                     /note="GRAIL"
                     /evidence=not_experimental
     exon            9223..9282
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   9191..9229
                     /rpt_family="Tigger2a"
     exon            9096..9528
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   10179..10635
                     /note="Terminal_exon"
     repeat_region   10824..10879
                     /rpt_family="MER31B"
     repeat_region   10882..11145
                     /rpt_type=tandem
     CDS             join(8470..8690,9077..9127,9266..9336,9948..10012,
                     10957..11014,12096..12164,12933..13064,14902..14960,
                     16609..16667,16745..16818,17370..17482,18671..18766,
                     21539..21677,21757..21842,23553..23671,24619..24715,
                     27291..27429,27763..27801,27930..28039,28733..28853,
                     29101..29179,29710..29854,30406..30624,31242..31277)
                     /gene="Elac2"
                     /codon_start=1
                     /note="25 exon form"
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAK29420.1"
                     /db_xref="GI:13540342"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPKDPLRHLRTREKRGPG
                     GPNTVYLQVVAAGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERCGKQOPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNPLVLKAKELGLPVGTRAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDEGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVELGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCLSLTAVFVSHLIADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     ETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAPGDIEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"
     repeat_region   complement(333..592)
                     /protein_id="AAK29421.1"
                     /db_xref="GI:13540343"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPKDPLRHLRTREKRGPG
                     GPNTVYLQVVAAGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERCGKQOPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNPLVLKAKELGLPVGTRAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDEGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVELGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCLSLTAVFVSHLIADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     ETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAPGDIEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"

BASE COUNT     8641 a    8030 c    8142 g    9780 t
ORIGIN

Qy        891 gggaaagatgcaccctcctgatacatgaagccac 925
              |||||||||||||||||||||||||||||||||||
Db      29100 GGGAAAGATGCACCCTCCTGATACATGAAGCCAC 29134

Query Match     3.58%;  Score 35;  DB 94;  Length 34593;
Best Local Similarity   100.0%;  Pred. No. 1.3e-08;
Matches 35;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
repeat_region    1138..11339
                 /rpt_family="MER31-internal"
repeat_region    11351..11994
                 /rpt_family="MER31-internal"
repeat_region    11563..11606
                 /rpt_family="MER31-internal"
repeat_region    12029..12223
                 /rpt_type=tandem
repeat_region    12122..12742
                 /rpt_family="MER57-internal"
repeat_region    12457..13937
                 /rpt_family="MER31-internal"
repeat_region    12637..13415
                 /rpt_family="MER4-internal"
repeat_region    12749..14506
                 /rpt_family="MER65-internal"
repeat_region    13011..13027
                 /rpt_family="MER31-internal"
exon             complement(13415..13937)
                 /note="XPOUND"
                 /evidence=not_experimental
exon             14127..14173
                 /note="XPOUND"
                 /evidence=not_experimental
exon             14239..14331
                 /note="GRAIL"
                 /evidence=not_experimental
exon             14439..14470
                 /note="XPOUND"
                 /evidence=not_experimental
repeat_region    14513..14985
                 /rpt_family="MER31B"
repeat_region    14986..15205
                 /rpt_family="LIPA13"
repeat_region    15583..15624
                 /rpt_family="MER31-internal"
repeat_region    complement(15594..16226)
                 /rpt_type=tandem
repeat_region    complement(15778..16081)
                 /rpt_family="AluYa8"
repeat_region    complement(16414..16776)
                 /rpt_family="AluSc"
repeat_region    complement(17761..18189)
                 /rpt_family="THE1B"
repeat_region    complement(18043..18271)
                 /rpt_family="LIPA13"
repeat_region    18286..18554
                 /rpt_family="L1"
exon             18291..18397
                 /rpt_family="LIPA10"
exon             18291..18397
                 /note="MEF"
                 /evidence=not_experimental
exon             18291..18397
                 /note="GRAIL"
                 /evidence=not_experimental
repeat_region    20013..20014
                 /rpt_family="MER83"
repeat_region    complement(19781..20183)
                 /rpt_family="MLT1F"
repeat_region    complement(18661..19071)
                 /rpt_family="MIR"
exon             21802..21906
                 /note="GRAIL"
                 /evidence=not_experimental
repeat_region    complement(21158..21279)
                 /rpt_family="L2a"
repeat_region    22911..23014
                 /rpt_family="AluJb"
repeat_region    complement(22020..22325)
                 /rpt_family="MER5A"
repeat_region    23103..23196
                 /rpt_family="MLT1J"
exon             complement(23581..23725)
                 /note="GRAIL"
                 /evidence=not_experimental
```

```
exon             23583..23663
                 /note="GRAIL"
                 /evidence=not_experimental
exon             24026..24185
                 /note="GRAIL"
                 /evidence=not_experimental
exon             complement(24093..24257)
                 /note="Terminal_exon"
                 /evidence=not_experimental
exon             24114..24185
                 /note="XPOUND"
                 /evidence=not_experimental
exon             24201..24344
                 /note="GRAIL"
                 /evidence=not_experimental
repeat_region    complement(24884..24945)
                 /rpt_family="LIMA6"
exon             complement(25524..25621)
                 /note="XPOUND"
                 /evidence=not_experimental
repeat_region    25642..25699
                 /rpt_family="L2"
repeat_region    25837..25924
                 /rpt_type=tandem
repeat_region    26272..27029
                 /rpt_family="LIPA7"
exon             26733..26881
                 /note="GRAIL"
                 /evidence=not_experimental
exon             27656..27783
                 /note="GRAIL"
                 /evidence=not_experimental
repeat_region    complement(27765..28027)
                 /rpt_family="LTR37B"
repeat_region    28150..28450
                 /rpt_family="MER33"
repeat_region    28452..29399
                 /rpt_family="LIMA5"
repeat_region    28897..29188
                 /rpt_family="AluSx"
repeat_region    29402..29593
                 /rpt_family="Charlie5"
```

```
Qy        289   aagaagtcagtaccacagaaatca 311
                |||||||||||||||||||||||
Db      39697   AAGAAGTCAGTACCACAGAAATCA 39719

Query Match             2.3%;    Score 23;   DB 89;    Length 105490;
Best Local Similarity  100.0%;   Pred. No. 0.14;
Matches   23;  Conservative  0;  Mismatches  0;   Indels    0;   Gaps    0;
```

COMMENT   On Jun 16, 2000 this sequence version replaced gi:7717162.
            ------------------ Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----------------- Project Information -----------------
            Center project name: H_NH0104D16
            ------------------ Summary Statistics ------------------
            Sequencing vector: M13; 49%
            Sequencing vector: plasmid; 51%
            Chemistry: Dye-primer ET; 49% of reads
            Chemistry: Dye-terminator Big Dye; 51% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182645 bases at least Q40
            Consensus quality: 185230 bases at least Q30
            Consensus quality: 186612 bases at least Q20
            Insert size: 199000; agarose-fp
            Insert size: 189205; sum-of-contigs
            Quality coverage: 4.52 in Q20 bases; agarose-fp
            Quality coverage: 4.79 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   2064: contig of 2064 bp in length
            *   2065   2164: gap of unknown length
            *   2165   3805: contig of 1641 bp in length
            *   3806   3905: gap of unknown length
            *   3906   8649: contig of 4744 bp in length
            *   8650   8749: gap of unknown length
            *   8750   13321: contig of 4572 bp in length
            *   13322   13421: gap of unknown length
            *   13422   19509: contig of 6088 bp in length
            *   19510   19609: gap of unknown length
            *   19610   24715: contig of 5106 bp in length
            *   24716   24815: gap of unknown length
            *   24816   30643: contig of 5828 bp in length
            *   30644   30743: gap of unknown length
            *   30744   40713: contig of 9970 bp in length
            *   40714   40813: gap of unknown length
            *   40814   48495: contig of 7682 bp in length
            *   48496   48595: gap of unknown length
            *   48596   57063: contig of 8468 bp in length
            *   57064   57163: gap of unknown length
            *   57164   67250: contig of 10087 bp in length
            *   67251   67350: gap of unknown length
            *   67351   82559: contig of 15209 bp in length
            *   82560   82659: gap of unknown length
            *   82660   102558: contig of 19899 bp in length
            *   102559   102658: gap of unknown length
            *   102659   124552: contig of 21894 bp in length
            *   124553   124652: gap of unknown length
            *   124653   152228: contig of 27576 bp in length
            *   152229   152328: gap of unknown length
            *   152329   190705: contig of 38377 bp in length

FEATURES   Location/Qualifiers
     source   1..190705
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-104D16"
     misc_feature   1..2064
                     /note="assembly_name:contig7"
     misc_feature   2165..3805
                     /note="assembly_name:Contig8
                     clone_end:SP6
                     vector_side:left"

     misc_feature   3906..8649
                     /note="assembly_name:Contig9"
     misc_feature   8750..13321
                     /note="assembly_name:Contig10"
     misc_feature   13422..19509
                     /note="assembly_name:Contig11"
     misc_feature   19610..24715
                     /note="assembly_name:Contig12"
     misc_feature   24816..30643
                     /note="assembly_name:Contig13"
     misc_feature   30744..40713
                     /note="assembly_name:Contig14"
     misc_feature   40814..48495
                     /note="assembly_name:Contig15"
     misc_feature   48596..57063
                     /note="assembly_name:Contig16"
     misc_feature   57164..67250
                     /note="assembly_name:Contig17"
     misc_feature   67351..82559
                     /note="assembly_name:Contig18"
     misc_feature   82660..102558
                     /note="assembly_name:Contig19"
     misc_feature   102659..124552
                     /note="assembly_name:Contig20"
     misc_feature   124653..152228
                     /note="assembly_name:Contig21"
     misc_feature   152329..190705
                     /note="assembly_name:Contig22"
BASE COUNT   55947 a   36657 c   35874 g   60716 t   1511 others
ORIGIN

RESULT   13
AL139399/c
DEFINITION Human DNA sequence from clone RP11-574A21 on chromosome
            Xq21.33-22.3, complete sequence.
LOCUS   AL139399   64151 bp   DNA   PRI   21-DEC-2000
ACCESSION   AL139399
VERSION   AL139399.9 GI:11989998
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 64151)
  AUTHORS   Howden,P.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT   On Dec 22, 2000 this sequence version replaced gi:11878435.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:

Qy   289 aagaagtcagtaccagaaatca 311
                |||||||||||||||||||||||
Db   161565 AAGAAGTCAGTACCAGAAATCA 161587

Query Match   2.3%; Score 23; DB 65; Length 190705;
Best Local Similarity   100.0%; Pred. No. 0.14;
Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

FEATURES
     source
                    Emi; EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
                    on the WORMPEP database can be found at
                    http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
                    was generated from part of bacterial clone contigs of human
                    chromosome X, constructed by the Sanger Centre Chromosome X Mapping
                    Group. Further information can be found at
                    http://www.sanger.ac.uk/HGP/chrX
                    RP11-574A21 is from the library RPCI-11.2 constructed by the group
                    of Pieter de Jong. For further details see
                    http://www.chori.org/bacpac/home.htm
                    VECTOR: pBACe3.6
                    IMPORTANT: This sequence is not the entire insert of clone
                    RP11-574A21 It may be shorter because we sequence overlapping
                    sections only once, except for a 100 base overlap.
                    The true left end of clone RP11-274M8 is at 64052 in this sequence.
                    The true right end of clone RP11-267M20 is at 100 in this sequence.
                    Location/Qualifiers
                    1. 64151
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="X"
                    /map="q21.33-22.3"
                    /clone="RP11-574A21"
                    /clone_lib="RPCI-11.2"

BASE COUNT   20073 a   11574 c   12196 g   20308 t
ORIGIN

Qy   212 ctgcagcttcccaacttccag 232
           ||||||||||||||||||||||
Db   24576 CTGCAGCTTCCCAACTTCCAG 24556

Query Match   2.1%;   Score 21;   DB 90;   Length 64151;
Best Local Similarity   100.0%;   Pred. No. 2;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 14
LOCUS   AB046207   1176 bp   mRNA   VRT   11-APR-2001
DEFINITION Anguilla japonica mRNA for 28kDa-1d apolipoprotein, complete cds.
ACCESSION   AB046207
VERSION   AB046207.1 GI:13591607
KEYWORDS   .
SOURCE   Anguilla japonica
  ORGANISM Anguilla japonica
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Anguilliformes;
            Anguilloidei; Anguillidae; Anguilla.
REFERENCE   1 (sites)
  AUTHORS   Kondo,H., Kawazoe,I., Nakaya,M., Kikuchi,K., Aida,K. and Watabe,S.
  TITLE   The novel sequences of major plasma apolipoproteins in the eel
  JOURNAL   Biochim. Biophys. Acta 1531 (1-2), 132-142 (2001)
  MEDLINE   21175592
REFERENCE   2 (bases 1 to 1176)
  AUTHORS   Watabe,S., Kondo,H., Kawazoe,I., Nakaya,M., Kikuchi,K. and Aida,K.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2000) Shugo Watabe, The University of Tokyo,
            Graduate School of Agricultural and Life Sciences; Yayoi 1-1-1,
            Bunkyo, Tokyo 113-8657, Japan
            (E-mail:awatabe@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-7520,
            Fax:81-3-5841-8166)
FEATURES   Location/Qualifiers
     source   1. 1176
                     /organism="Anguilla japonica"
                     /db_xref="taxon:7937"
                     /tissue_type="liver"
     CDS   18. 797
                     /codon_start=1
                     /product="28kDa-1d apolipoprotein"
                     /protein_id="BAB40964.1"
                     /db_xref="GI:13591608"
                     /translation="MKFIVVIALVSFTGCQANVLRSNEPTFPQLELVKDAFYFVKAS
                     QTAQGALKTFGESELAQQVNAKIKESVEVQVQYRRIVQEQVIVISDELHKKLSEHAEQ
                     LSESLQPDINEVRVQLEPLAEKLSANIQQOMOKVRVPLDPYTESLDIRALERALRMH
                     RRLFMSVEQLLSQLLEHLQDLQDLGPGSTEELKRVEESVQEYRRFVLPLENIIGNEILNK
                     LPMFFRRVYPVVDGLREKLDPYQIQGLESQLTALWEESSANSA"

BASE COUNT   333 a   262 c   286 g   295 t
ORIGIN

Qy   234 agagcgtgcaggagtacagg 253
           ||||||||||||||||||||||
Db   601 AGAGCGTGCAGGAGTACAGG 620

Query Match   2.0%;   Score 20;   DB 8;   Length 1176;
Best Local Similarity   100.0%;   Pred. No. 7.9;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 15
LOCUS   ATU23796   1573 bp   mRNA   PLN   21-JUL-1995
DEFINITION Arabidopsis thaliana IAA-amino acid hydrolase homolog ILL2 (ILL2)
            mRNA, complete cds.
ACCESSION   U23796
VERSION   U23796.1 GI:902790
KEYWORDS   .
SOURCE   Arabidopsis thaliana
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1573)
  AUTHORS   Bartel,B. and Fink,G.R.
  TITLE   ILR1, an amidohydrolase that releases active indole-3-acetic acid
            from conjugates
  JOURNAL   Science 268 (5218), 1745-1748 (1995)
  MEDLINE   95312867
REFERENCE   2 (bases 1 to 1573)
  AUTHORS   Bartel,B. and Fink,G.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-MAR-1995) Bonnie Bartel, Whitehead Institute, 9
            Cambridge Center, Cambridge, MA 02142, USA
COMMENT   On Jul 21, 1995 this sequence version replaced gi:887788.
FEATURES   Location/Qualifiers
     source   1. 1573
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /chromosome="5"
     CDS   1. 1320
                     /gene="ILL2"
                     /note="ILR1-like gene; IAA-amino acid hydrolase homolog"
                     /codon_start=1
                     /product="ILL2"
                     /protein_id="AAC49016.1"
                     /db_xref="GI:902791"
                     /translation="MALNKLLSLTFQLLLFLLSVSSESPWIAEDISQIQTKLLEFAKS
                     PEVFDWMVKIRRKIHENPELGYEELETSKIIRSELELEIGIRYRPVAITGVIGYIGTG
                     EPPEVALRADMDALPIOEGVEWEHKSKIPGKMHACHDGHYTMLLGAAILHEHRHL
                     QGTVVLIFQPAEEGLSGAKKMREEGALKNVEAIFGIHLSARIPFGKAASRAGSFLAGA
                     GVFEAVITGKGGHAAIPHHTIDPVMAASSIVLSLQQLVSRETDPLDSKVVTVSKVNGG
                     NAFNVIPDSITIGGTIRAFTGFTIQLCQORVKEVITKQAAVHRCNASVNLTGNGRRPP
                     TVNNRDLYKQFKKVRDLLGQEAFVEAAPVWGSEDFSYFAETIPGHFSLLGMQDETNG
                     YASSHSPLYRINEDVLPYGAAIHASMAVQYLKEKASKGSVSGFHEEL"
     gene   75. 1320
                     /gene="ILL2"

BASE COUNT   452 a   332 c   363 g   426 t
ORIGIN

Query Match   2.0%;   Score 20;   DB 14;   Length 1573;
Best Local Similarity   100.0%;   Pred. No. 7.9;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 16

```
Qy    565 ctttgggaaagccgcttcac 584
          ||||||||||||||||||||
Db    609 CTTTGGGAAAGCCGCTTCAC 628
```

Db   AF047031

```
LOCUS       AF047031    4918 bp    DNA          PLN       03-MAY-1999
DEFINITION  Arabidopsis thaliana IAA-amino acid hydrolase (ILL1 and ILL2)
            genes, complete cds.
ACCESSION   AF047031
VERSION     AF047031.1  GI:2921028
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 4918)
  AUTHORS   Bartel,B. and Fink,G.R.
  TITLE     ILR1, an amidohydrolase that releases active indole-3-acetic acid
            from conjugates
  JOURNAL   Science 268 (5218), 1745-1748 (1995)
  MEDLINE   95312867
REFERENCE   2 (bases 1 to 4918)
  AUTHORS   Davies,R.T., Goetz,D.H., Lasswell,J., Anderson,M.N. and Bartel,B.
  TITLE     IAR3 encodes an auxin conjugate hydrolase from Arabidopsis
  JOURNAL   Plant Cell 11 (3), 365-376 (1999)
  MEDLINE   99172052
REFERENCE   3 (bases 1 to 4918)
  AUTHORS   Bartel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-1998) Biochemistry and Cell Biology, Rice
            University, 6100 S. Main St., Houston, TX 77005, USA
FEATURES             Location/Qualifiers
     source          1..4918
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /map="between nga129 and LFY3"
                     /chromosome="5"
     gene            join(<143..496,605..730,805..1119,1201..1323,1408..>1806)
                     /gene="ILL1"
     mRNA            join(<143..496,605..730,805..1119,1201..1323,1408..>1806)
                     /gene="ILL1"
                     /product="IAA-amino acid hydrolase"
     gene            <143..>1806
                     /gene="ILL1"
                     /note="ILR1-like gene 1"
     CDS             join(143..496,605..730,805..1119,1201..1323,1408..1806)
                     /gene="ILL1"
                     /function="auxin conjugate hydrolase"
                     /codon_start=1
                     /product="IAA-amino acid hydrolase"
                     /protein_id="AAC04865.1"
                     /db_xref="GI:2921829"
                     /translation="MALNNFITFQLLLLLLRVSSESPWIVAGDVSRIPINFLELAKSP
                     EVFDSMYIRRKIHENPELGYEEFETSKFIRSELDLIGVKYRFPVAITGICYIGTGE
                     PPFVALRADMDALPIQEAVEWEHKSKNPGKMHACGHDGHVAMLLGAAKILQQHROLHLQ
                     GTVVLIFQPAEEGLSGAKMMREEGALKNVEAIFGIHLSPRTEPGKAASLAGSEMAGAG
                     AFNVIPDSITIGCTLRAFTGTFTQLOERIKEIITKQAAVHRCNASVNLAPNGNQPMPPT
                     VNNMDLYKFKFKFKVVRDLLGQEAFVEAVPEMGSEDFSYFAETIPGHFSLLGMQDETGY
                     ASSHSPHYRINEDVLPYGAAIHATMAVQYLKDKASKGSSVSGFHDEL"
     mRNA            join(<2596..2952,3121..3246,3323..3637,3718..3840,
                     3931..>4329)
                     /gene="ILL2"
                     /product="IAA-amino acid hydrolase"
     gene            <2596..>4329
                     /gene="ILL2"
                     /note="ILR1-like gene 2"
     CDS             join(2596..2952,3121..3246,3323..3637,3718..3840,
                     3931..4329)
                     /gene="ILL2"
                     /function="auxin conjugate hydrolase"
                     /codon_start=1
                     /product="IAA-amino acid hydrolase"
                     /protein_id="AAC04866.1"
                     /db_xref="GI:2921830"
                     /translation="MALNKLLSLIFQLLLFLLSVSSESPWIAEDTSQIQTKLLEFAKS
                     PEVFDWMVKIRRKIHENPELGYEELETSKLIRSELELGIKYRYPVAITGVGIVIGTG
                     EPPFVALRADMDALPIQEGVEWEHKSKIAGKMHACGHDGHVTMLLGAAKILHEHRHL
                     QGTVVLIFQPAEEGLSGAKMKREEGALKNVEAIFGIHLSARIPFGKAASRAGSFLAGA
                     GVFEAVIGCKGIHAAIPQHTIDPVVAASSIVLSLQQLVSREIDPLDSKVVTVSKVNGG
                     NAFNVIPDSITIGGTLRAFTGTFTGFTQLQORVKEVITKQAAVHRCNASVNLTPNGREPNPP
                     TVNNKDLYKFQFKKVVRDLLGQEAFVEAAPVMGSEDFSYFAETIPGHFSLLGMQDETNG
                     YASSHSPLYRINEDVLPYGAAIHASMAVQYLKEKASKGSSVSGFHEEL"
     variation       3082..3083
                     /gene="ILL2"
                     /note="6 base insertion in ecotype Landsberg erecta"
                     /replace="tgaatt"
BASE COUNT     1422 a    991 c    997 g    1508 t
ORIGIN
```

RESULT 17

```
Qy    565 ctttgggaaagccgcttcac 584
          ||||||||||||||||||||
Db   3448 CTTTGGGAAAGCCGCTTCAC 3467
```

Query Match   2.0%;   Score 20;   DB 12;   Length 4918;
Best Local Similarity   100.0%;   Pred. NO. 7.9;
Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Db   3448 CTTTGGGAAAGCCGCTTCAC 3467

```
LOCUS       AB014516    6718 bp    mRNA          PRI       06-FEB-1999
DEFINITION  Homo sapiens mRNA for KIAA0616 protein, partial cds.
ACCESSION   AB014516
VERSION     AB014516.1  GI:3327045
KEYWORDS    .
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescript
            SK plus clone:HG03623.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 6718)
  AUTHORS   Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) to the DDBJ/EMBL/GenBank databases. Osamu
            Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
REFERENCE   2 (sites)
  AUTHORS   Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
            Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. X.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (3), 169-176 (1998)
  MEDLINE   98403880
FEATURES             Location/Qualifiers
     source          1..6718
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HG03623"
                     /clone_lib="pBluescript II SK plus"
                     /dev_stage="adult"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..1906
                     /gene="KIAA0616"
     CDS             <1..1906
                     /gene="KIAA0616"
                     /codon_start=2
                     /product="KIAA0616 protein"
```

```
                                    /protein_id="BAA31591.1"
                                    /db_xref="GI:3327046"
                                    /translation="HNQKQAEETAAFEEVMKDLSLTRAARLQLQKSQYLQLGPSRGQY
                                    YGGSLPNVNQIGSGTMDLPFQPSGFIGEAALAAAPVSLTPQSSGLDTSRITRHHLVD
                                    RVYERGRLGSPHRRELSVDHGRQADSCPYGTMYLSPPADTSWRRTNSDSALHQSTM
                                    TPTQPESFSSGSSQDVHQKRVLLTVPGMEETTSEADKNLSKQAWDTKKTGSRPKSCEV
                                    PGINIPPSAQDENTTALIPATHNTGGSILPDLINIHFPSPLPTLPDEEPTPALSSSS
                                    STGNLAANLIHLGIGGAGOGMSTPGSSPQHRPAGVSPLSLSTEARRQCQASPTLSPLSP
                                    ITQAVAMDALSLEQOLPYAFFTQAGSQOPPPQPQOPPPPPPPPASQQPPPQOPAPVRL
                                    PPGGPLLPSASLITRGPQOPPLAVTVPSSLPQSPENPGQPSMGIDITASAPALQQYRTS
                                    AGSPANQSPTSPVSNQGFSPGSSPQHTSTLGSVFGDAYYEQOMARQANALSHQLEQF
                                    NMMENAISSSSLYSPSSTLNYSQAAMMGLTGSHSGLSPDSQOLGYASHSGIPNIILIVT
                                    GESPPSLSKELTSSLAGVGVDVSFDSDSQOFPLDELKIDPLITLDGLIHMLNDPDMVLADPA
                                    TEDTFRMDRL"

BASE COUNT    1208 a   2237 c   1930 g   1343 t
ORIGIN

RESULT 18
AC003107
LOCUS       AC003107    46275 bp    DNA          PRI       18-NOV-1997
DEFINITION  Human DNA from chromosome 19-specific cosmid R30064 containing the
COMP gene, genomic sequence, complete sequence.
ACCESSION   AC003107
VERSION     AC003107.1   GI:2623749
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46275)
  AUTHORS   Lanerdin,J.E., McCready,P.M., Adamson,A.W., Burkhart-Schultz,K.,
            Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J.,
            Danganan,L., Christensen,M., Bruce,R., Quan,G., Montgomery,M.,
            Ow,D., Kobayashi,A., Olsen,A.O. and Carrano,A.V.
  TITLE     Sequence analysis of an ~1 Mb region containing the MEF2B gene in
            19p12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46275)
  AUTHORS   Lanerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) Human Genome Center, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Cosmid R30064 overlaps cosmid F19807 to the left and cosmid R32269
            to the right.

FEATURES             Location/Qualifiers
     source          1..46275
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R30064"
                     /chromosome="19"
                     /map="19p12 between UBA52 and D19S451"
                     /cell_line="5H2-B"
                     /clone_lib="LL19NCO3 cosmid library"
                     /note="LL19NCO3 R chromosome 19 cosmid library"
                     flow-sorted chromosomes from hybrid 5H2-B, which carries
                     chromosome 19 as its only human chromosome."
     misc_feature    817..1037
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 1, quality: excellent, score: 93.000"
     repeat_region   complement(3847..3899)
                     /rpt_family="MER7"
```

```
misc_feature    /note="predicted exon, program: grail2exons_human_1.3,
18533..18757       frame: 2, quality: excellent, score: 100.000"
misc_feature    /note="DDS similarity to AA252605 zs14b02.s1 NCI_CGAP_GCB1
20682..22267       Homo sapiens cDNA clone 685131 3' end. Score: 424
                   Identity: 225/225 (100%)."
                /note="DDS similarity to overlapping ESTs:
                (20682..21075) AA350958 EST58512 Infant brain Homo sapiens
                cDNA 5' end; Score: 776 Identity: 391/394 (99%).
                (20863..21259) R87396 ym88h07.r1 Homo sapiens cDNA clone
                166045 5'; Score: 748 Identity: 391/401 (97%).
                (20874..21249) AA349930 EST57030 Infant brain Homo sapiens
                cDNA 5' end; Score: 748 Identity: 375/376 (99%).
                (21125..21574) R19558 yg26h07.r1 Homo sapiens cDNA clone
                33726 5'; Score: 811 Identity: 440/457 (96%).
                (21589..22030) H30788 yo79f06.r1 Homo sapiens cDNA clone
                184163 5' similar to gb:M33326 NONSPECIFIC CROSS-REACTING
                ANTIGEN NCA-95 (HUMAN);contains Alu repetitive element.
                Score: 817 Identity: 433/443 (97%).
                (22007..22267) AA322326 EST24948 Cerebellum II Homo
                sapiens cDNA 5' end; Score: 479 Identity: 252/262 (96%).
                "
repeat_region   complement(21512..21626)
                /rpt_family="Alu"
misc_feature    22506..22980
                /note="DDS similarity to R88263 ym90h09.s1 Homo sapiens
                cDNA clone 166241 3'. Score: 837 Identity: 455/478
                (95%).-Other overlapping matches:
                (23013..22497) H05926 yl71f07.s1 Homo sapiens cDNA clone
                43353 3'; Score: 872 Identity: 497/527 (94%).
                (22971..22534) AA446916 zw85a01.s1 Soares total fetus
                Nb2HF8 9w Homo sapiens cDNA clone 783720 3'; Score: 857
                Identity: 436/437 (99%).
                (22971..22561)W95490 ze02g07.s1 Soares fetal heart NbHH19W
                Homo sapiens cDNA clone 357852 3';
                Score: 774 Identity: 399/411 (97%).
                (22566..22949) W95528 ze02g07.r1 Soares fetal heart
                NbHH19W Homo sapiens cDNA clone 357852 5';Score: 674
                Identity: 366/382 (95%).
                (22996..22634) T33668 EST58674 Homo sapiens cDNA 3' end
                similar to None. .; Score: 710 Identity: 359/363 (98%)."
misc_feature    complement(23576..23770)
                /note="DDS similarity to N94385 zb76g06.s1 Soares
                senescent fibroblasts NbHSF Homo sapiens cDNA clone 309515
                3'.
                similar to gb:219585 THROMBOSPONDIN 4 PRECURSOR (HUMAN);
                (194..1); 92% Identity.-Other overlapping matches:
                (23724..23770) predicted exon, program:
                grail2exons_human_1.3, frame: 1, quality: good, score:
                74.000-(23770..23664) DDS similarity to W36954 mb82e10.r1
                Soares mouse p3NMF19.5 Mus musculus cDNA clone 335946 5';
                similar to
                gb:219585 THROMBOSPONDIN 4 PRECURSOR (HUMAN); gb:M86620
                Mouse thrombospondin 3 (MOUSE) (307..423); 63% identity."
gene            complement(23683..32077)
                /gene="COMP"
                /note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR
                (COMP)"
CDS             complement(join(23724..23770,23683..24002,25000..25172,
                25705..25901,26307..26355,26482..26774,26955,
                27048..27100,27341..27459,28299..28458,29020..29127,
                29218..29322,29400..29558,29647..29721,29968..30105,
                30750..30922,31370..31421,31658..31743,31999..32077))
                /gene="COMP"
                /note="CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR
                (COMP)"
                /codon_start=1
                /product="COMP_HUMAN"
                /protein_id="AAB86501.1"
                /db_xref="GI:2623750"
                /translation="MVPDTACVLLTLAALGASGQGQGSPLGSDLGPQMLRELQETNAA
                LQDVRELLRQQVREITFLKNTVMECDACGMQQSVRIGLPSVRPLLHCAPGCFPGVAC"
```

```
RESULT 19
LOCUS       AB007727      79590 bp    DNA             PLN       27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MXC9.
ACCESSION   AB007727 BA000015
VERSION     AB007727.1  GI:2696018
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MXC9.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (sites)
  AUTHORS   Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. III.
            Sequence features of the regions of 1,191,918 bp covered by
            seventeen physically assigned P1 clones
  JOURNAL   DNA Res. 4(6), 401-414 (1997)
  MEDLINE   98162728
REFERENCE   2 (bases 1 to 79590)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MXC9
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers. Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremln1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
```

```
Query Match             2.0%;   Score 20;  DB 85;  Length 46275;
Best Local Similarity  100.0%;  Pred. No. 7.9;
Matches   20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      462  gggtcctggggcaccctggct 481
             |||||||||||||||||||||
Db    20016  GGGTCCTGGGGCACCCTGGCT 20035
```

```
IQTESGARCGPCPAGFTGNGSHCTDVNECNAHCFPRVRCINTSPGFRCEACPPGYSG
PTHQGVGLAFAKANKQVCTDINECETGQHNCVPNSVCINTRGSPGCGPCQPGFVGDQA
SGCQRRAQRPCPDGSPSECHEHADCVLERDGSRSCVCAVGWAGNGILCGRDTDLDGFP
DEKLRCPERQCRKDNCVTVPNSGQEDVRDGIGDACDPDADDGVPNEKDNCPLVRNP
DQRNTEDRWGDACDNCRSQKNDDQKTDQDGRGDACDDDIDGDRIRNQADNCPRVPN
SDQKDSDGDGIGDACDNCPQKSNPDQADVDHDFEVGDACDSDQDGDGDADRDGVPN
VPNSAQEDSDHDGQDGDACDDDRDNCRLVPNPGQEDADRDGVGDVCDDDF
DADKVVDKITDVCENAEVTLTDRRAFQTYVLLDDEPEGDAQIDPNWVLNQGREIVQTMNS
DPGLAVGITAFNGVDFEGTFHVNTVTDDDYAGFIFGYQDSSSFYVVMMKQMEQTYWQA
NPFRAVAEPGIQLKAVKSSTGPEQLRNALWHTGDTESQVRLLWKDPRNVGWKDKKSY
```

CDS
```
/clone="MXC9"
/clone_lib="Mitsui P1"
complement(3423..3875)
/note="contains similarity to S1 self-incompatibility
protein
gene_id:MXC9.1"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10024.1"
/db_xref="GI:9759373"
/translation="MKNILKTQVHVVVIYLLIKIAFSQVKFDFDVNWSTSKMVRITNR
LGDGILTNLHCKSADDDLGLKILAPNGSWSFKFRPSIVGTTLFYCHFTWPGQSKRFDI
YDDDRDGVRSHSICINCIWDISIQGPCMFSESDHAFNICYDWNGNLRT"
```

CDS
```
complement(5454..5903)
/note="gene_id:MXC9.2
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10025.1"
/db_xref="GI:9759374"
/translation="MLKTQVHVVVIYLLIQIAFSQVKPGSDLKWSTLKSVVTINRLG
DGSTLKLHCKSADDDLGLKILAPNGSWSFKPRPSIVPGVTLFFCHFTWPGQSKWFNIY
DDDRDGVRMGIPCICIWDIGKYGPCRFSEIDDAFNICYDWNGNRRS"
```

CDS
```
complement(join(6686..6926,7008..7210,7299..7592,
7870..7993,8085..9427))
/note="embl:CAB87679.1
gene_id:MXC9.3
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/db_xref="GI:9759375"
/translation="MAEQKSSNGGGGGGGGGDVVINVPVEEASRRSKEMASPESEKGVPFS
KSPSPEISKLVGSPNKPPRAPNQNNVGLTQRKSFARSVYSKRSPRFVDPSCPVDTSIL
EEEVRROLGAGFSFSRASPNNKSNRSVGSPAPVTPSKVVVEKDEDEIYKKVKLNREM
RSKISTLALLIESAFPVVILSALVASLTINVLKHHTFWGLEVWKNCVLVWVIFSGMLVT
NWFMRLIVFLPIETNFLLRRKVLYFVHGLKSVQVFIWLCILVAMILLNHDVKRSPA
ATKVLKCITRYLISLITGAFFWLVKTLLLKILAANFNVNNFFDRIQDSVPHQYVLQTL
SGLPLMEEAERVGREPSTGHLSFATVVKGTVEKRKVIDMGKVHKMKREXVSAWTMRV
LMEAVRTSGLSTISDTLDETAYGEQEQADREITSEMEALAAAYHVFRNVAQFFENYI
EEEDLLRFMIKEEVDLVFLPDGAAETGRITRKAFTEWVVKVYTSRRALAHSLNDTKT
AVKQLNKLVTAILMVVTVIWLLLEVATTKVLFFSTQLVALAFIIGSTCKNLFESI
VFVPWHPIPVDGRCVVDGVAMLVEEMNLLTTVFLKLRERTAEYLEONPQHWAPVHSVVVKEIENMKLK
RSPNMGETVEFSISFSTPVSKIAHLKERTAEYLEONPQHWAPVHSVVVKEIENMKLK
MALYSDHTITFQENRERNLRRTELSLAIKRMLEDLHIDYTLLPQDINLIKKN"
```

CDS
```
complement(13439..17695)
/note="gene_id:MXC9.4"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="retroelement pol polyprotein-like"
join(18207..18234,18300..19381)
```

CDS
```
join(18207..18234,18300..19381)
/note="contains similarity to NPK1-related protein kinase
gene_id:MXC9.5"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10027.1"
/db_xref="GI:9759376"
/translation="MIENDSPQHPSLFFTDFTRANSHVRVLHHKFLGKGVYGSVDLI
RYTKTDGSSLQAAVKTSYAEDLEEYDALKREIQILSELKGYPNIVICYGDDLEEDFNE
HGHKYKYLLLEYANEGSLSSFMENYPDRLKLPDPMIRDFTRMILEGLVSMIHSGVYHCD
LKSDNLIIFSRKDSASCELKIFDFGNCRQVGEVPDHWKSDYPYVCGTPESPFDGVAKKT
LDLWSLGCLVLKIYTGEQPWERVISVDFVNFELSDGEAPNIPEYVPCDAREFIETCFAR
EHERGTAASELLLHFPLCGOKOPLLVKLRIKQASKITMGIREKPLTLKMFLSKPPQL
KKLLSKLLRLKKMIHPKPLELDSSPVYSSSDLCLL"
```

CDS
```
/note="gb|AAF19552.1
gene_id:MXC9.6
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10028.1"
/db_xref="GI:9759377"
```

```
/translation="MVTRLRLVSRSRYATVKFTDSVSACCGCRRLFSASTPPEPESQP
EQAPPNPVTGDEKLRNLVLLQQNRTETARGVLSSLLRSDSTPFASPKELFSAFSLS
SPSLKHDFSYLLLSVLLNESKMISEAADLFPALRNEGIYPSSDSLTLLLDHLVKIKQF
RVTINVFLNILESDFRPSKPMYGKAIQAAVKLSDVGKGLELFNRMKHDRIYSFVFIYN
VLIDGCKGKRMNDAEOLFDEMLARRLLPSLITTYNTLIDGYCKAGNPEKSPKVRERMK
ADHIEPSLITFNTLLKGLFKAGMEVADAENVLKEMKDLGFVPDAFTFSILFDGYSSNEK
AEAALGVYETAVDSGVKMNATTCSILLNALCKEGKIEKAEEILGREMAGLVPNEVIY
NTMIDGYCRKGDLVGARMKIEAMEKQGMKPDHIAYNCLIRRCELGEMENAEKEVNRM
KLKGVSPSVETIYNILLIGGYGRKYEFDKCFDILKEMEDNGTMPNVVSYGTLLINCLCKGS
KILEAQIVKRDMEDRGVSPKVRIYNLMIIDGCCSKGKIEDAFRFSKEMLKKGIELNIVT
YNTLIDGLSMTGKLSEAEDLLEISRGLKPDVFTVNSLLISGYGFAGNVQRCIALVEE
MKRSGIKPTLKTYHLLLSLCTKEGIELTERLFGEMSLKPDLLVYNGVLHCYAVHGDME
KAFNLOKOMIEKSIGIDKTTYNSLLIGOLKVGKICEVRSLIDEMNAREMEPEADTYNI
IVKGHCEVKDYMSAYVWNYREMQEKGFLLDVCIGNELVSGLKEEMWRSKEAEIVISEMNG
RMLGDVTVDEDLSATEKL"
join(22987..23072,23172..23288,23378..23453,23623..23752,
23874..23987,24073..24236)
/note="embl:CAB64729.1
gene_id:MXC9.7"
/codon_start=1
/evidence=not_experimental
/product="elongation factor 1B alpha-subunit"
/protein_id="BAB10029.1"
/db_xref="GI:9759378"
/translation="MAVTFSDLHTERGLKTLEEHLAGKTYISGDQLSVDDVKVYAAVL
ENPGDGFPNASKWYDSYASHLAKSFPGKADGVRVGGGVAPPSEAHPHTEEPAADGPGD
DDDIDLFADETEDEKKAAEERBAAAKKDTKKTKESGKSSVLLEVKPWDDETDMKKLEE
AVRSVQMPGLTWGASKLVPVGYGIKIKLTIMMTIVDDLVSVDNLIEDHLTSEPNNEYIO
SVDIVAPNKI"
```

CDS
```
complement(24733..26592)
/note="gene_id:MXC9.8
pir||T02643
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10030.1"
/db_xref="GI:9759379"
/translation="MSSASKMKSRDKKVMNDSOKTPSKASGSMGGSSVVGGYNPLLG
TFQTIESLSATICSSLHNNGRFRHIDESDSTGADCDSVSNNGSWGCGSEDHKEKYPST
SIAAKQEIIPGADNDRKDRMRLKNERKHQROKEKRAQELHERCCQILMSRKLEVHAQL
IISMGIPHERATYALMLNEGKIEESINWLFDDGGAVKKLDPTSGNLKLDISQELG
RILELETKYCTKQDVERAVVTAEGDIEKAEEALRRQKOQDQSTASRKPEDISDSTSVN
NSKVPSVLTSQNTLVQLOPNSGGHYPAGREEALDRHNLGIYPRGSSYISGSESNQSVNSL
ERIHMKLQWMKLOONAPLEENKRIPYQOTPLSRSTEETHYVTALGDGFKRLQOQDMRE
PVMVMQOOOOORSQSANTNVLPVSTMNASFTGAAAAGSGWHPANRSEAAQSNGYLP
SRSLPPNDLNSNLMTYQOLYTYQGQYQGQONNNNGHRMAGSSAPAVAPAASILGFSGFGSL
AGSSSSGLDWNFDGSLGHLDYNIDMSLDKGLACVRRSQOYVAASSPYSAASSPYEAH
MNGRTRMMTNGNGMGYVAMGVQBEAALVGNGREWTSPFEGKDLFSLSRQYVPPSL"
```

CDS
```
join(28348..28450,28599..28744,28826..28902,28902..29080,
29161..29241,29336..29392,29485..29566,29659..29708,
29854..29910,29996..30074,30190..30248,30406..30463,
30561..30658)
/note="contains similarity to transmembrane transport
protein
gene_id:MXC9.9"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10031.1"
/db_xref="GI:9759380"
/translation="MLMCFWVTLSAANRRLSHLPPIACSRGIDQEDEEKESRELLPH
KNDENATTSRRSSSSVDSGGGLKGYQQEETYKTSFKTVALCVGTAVAFGIGICKLEGVGK
ASEFFAGYILEQSLSVDNLFVFLVEKYFKVPLMYQNKVLIYGIAGAIVRFTLLILG
TATLQKFEAVNLLAAVLLYSSFKLFASEEDDTDLSDNFIVKTCQRFIPVTSSVDGNR
FFTKHDGILRKATPLLLTVAVIELSDIAFAVDKFGEGFHISTEASLGVVALSLSTGVL
SLYTLISEGMDELEYLQPSIAVVLGFIGVKMLDFFGFHISTEASLGVVALSLSTGVL
LSITNKSSDS"
complement(join(31799..31984,32321..32440))
/note="embl:CAA03929.1
gene_id:MXC9.10"
/codon_start=1
/evidence=not_experimental
/product="cystatin"
/protein_id="BAB10032.1"
```

```
Qry     760 atttggaagagttcagacc 779
            ||||||||||||||||||||
Db     4335 ATTTGGAAGAGTTCAGACC 4354
```

```
Query Match           2.0%;  Score 20;  DB 12;  Length 79590;
Best Local Similarity 100.0%;  Pred. No. 7.9;
Matches 20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
RESULT_20
AB013392
LOCUS       AB013392   84129 bp   DNA            PLN   27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MIK19.
ACCESSION   AB013392 BA000015
VERSION     AB013392.1  GI:3128139
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MIK19.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Miyajima,N.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. VI.
            Sequence features of the regions of 1,367,185 bp covered by 19
            physically assigned P1 and TAC clones
  JOURNAL   DNA Res. 5 (3), 203-216 (1998)
   MEDLINE   98403884
REFERENCE   2  (bases 1 to 84129)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MIK19
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MKN22 and the 3' clone is MPI10.
FEATURES             Location/Qualifiers
     source          1..84129
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MIK19"
                     /clone_lib="Mitsui P1"
                     /note="gene_id:MIK19.1"
     CDS             join(941..1855,1946..2080,2340..2609)
                     /pir||T06019
                     similar to unknown protein"
```

```
                     /db_xref="GI:9759381"
                     /translation="MADQQAGTIVGGVRDIDANANDLQVESLARFAVDEHNKNENLTL
                     EYRKLLGAKTQVVAGTMIHLTVEVADGETNKVYEAKVLEKAWENLKQLESFNLHDV"
```

```
CDS             complement(join(3056..3148,3239..3284,3384..3607,
                3708..3890,3970..4197,4294..4428,4528..4612,4749..4825))
                /note="gene_id:MIK19.2"
                /codon_start=1
                /product="protein kinase MEK1 homolog"
                /protein_id="BAB09874.1"
                /db_xref="GI:10176760"
                /evidence=not_experimental
                /translation="MRKLFFFLCTHVPKQRAKFTRIMESCQKKQKLETPIELSDMPDD
                LIFKIFSFLPFKEDLATRFISEYGKGLWNPDNAIFDDESLDMTCMSFVYGSLMSKD
                AQILDSLHLKLRKYIIASDINFLVQLGVNSMRELRIDFGELFLLPCCLSTCTTLKV
                LVLDHLNIMSVPGWFRLPSVETLQLSSVTGGSNYVPSLIRLCSVHERLVVDQNRKDL
                VNINVPFLRSLSIDNRRGHVPLGSPWINVPSLMFLNIKDFPQTIFESMPAMIKANIE
                VAHDQSQSVKFIESLTSTRHLCLCSPTSENPYPNGTKFSNLVHLKLCTCSAGWQNLLA
                CMLNDAPNLRSLTLKLRQKSDVNPKKVWEKFTVVPECLSTRLEILIKWRDYEGTEHEKD
                MVGYILANATFLQRAFSTKDRNQCDSRFSELQSMERVSEICEFVFD"
```

```
CDS             join(5726..5780,5997..7280,7375..7408,7494..7641)
                /note="gene_id:MIK19.3"
                /codon_start=1
                /evidence=not_experimental
                /product="beta-1,3-glucanase-like protein"
                /protein_id="BAB09876.1"
                /db_xref="GI:10176762"
                /translation="MARDFKLIFSISILLLLDCCYGKVGVCVGRSADDLPTPSKVV
                QLIQQHNIKYRIYDYNSQVLKAFGNTSIELMIGVPNSDLNAFSQSOSNVDTWLKNSV
                LPYYFTKIYIITVGAESTDDPHINASSFVVPAMONVLTALRKVELSRRIKVSTTLSL
                GILSRFPPSAGAFNSSYAYFLRPMLEFLAENKSPFMIDLYPYAYRDSPNVSLDYV
                LFESSSEVIDENTGLLYKNMFDAQVDALYALTALNFRTIKIMVTETGWPTKGSPREK
                AAASDNAEYYNSNIIRHVVTNQGTPAKPGEAMNVYIFSLFNENRKAGLLDSERNGLF
                YPDQTSVYQLDFTGKSNGFHSNSSGTNSSGGSNSWCIASSKASERDLKGALDWACGPG
                NVDCTAIQPSQPCFQPDTLVSHASFYFNSYFQQNRATDVACSFGGAGVKNKDPSYDK
                CIYIYAGGNTKATNATALTSSASFPRGNELLQWILKLCLMISLFFSLQTMNSQAL"
```

```
CDS             complement(join(8087..8221,8288..8425,8793..8915))
                /note="gene_id:MIK19.4"
                /codon_start=1
                /evidence=not_experimental
                /product="profilin-like protein"
                /protein_id="BAB09877.1"
                /db_xref="GI:10176763"
                /translation="MSWQTVYDDHLMCDVAGNRLTAAAILGQDGSVWAQSNNFPQVKP
                EEIQGIKDDFTTPGTLAPTGIFLGGNKYMVIQGEPNAVIRGKKGAGGVTIKKTTLALV
                FGIYDEPMTPGGCNMVVENLGEYLIESGL"
```

```
CDS             join(13604..13668,13748..14074,14151..14235)
                /note="gene_id:MIK19.5"
                /codon_start=1
                /evidence=not_experimental
                /product="retroelement pol polyprotein-like"
                /protein_id="BAB09878.1"
                /db_xref="GI:10176764"
                /translation="MTLYFIQKFVQEKIWQTYFQILSIVHHIQEDCGMYAGKKTPIM
                YEENAAYIAQLKDGYIKGDRFKHILPKFYTHDLORSGDVRVLQIRLNDNLADLFTKA
                LPTATFKKLVYRIGLRCLKNLDEYTHKGEQHFKYFSYGTIMVIRGSVMNKDGCPL"
```

```
CDS             join(14943..15139,15345..15429,15493..15582,15665..15758,
                15841..15923,16024..16182)
                /note="contains similarity to unknown protein
                gb|AAD15447.1
                gene_id:MIK19.6"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB09879.1"
                /db_xref="GI:10176765"
                /translation="MIDETEEDDTTQQRSRNDGVSKNKGKGFKGDKAKRALIGAGGR"
```

```
Qy    565  ctttgggaaagccgcttcac 584
           ||||||||||||||||||||
Db  32335  CTTTGGGAAAGCCGCTTCAC 32354

RESULT 21
LOCUS       ATF14F18     103638 bp    DNA    PLN    13-APR-2000
ATF14F18
DEFINITION  Arabidopsis thaliana DNA chromosome 5, BAC clone F14F18 (ESSA
            project).
ACCESSION   AL163812
VERSION     AL163812.1  GI:7573350
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 103638)
  AUTHORS   Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
            Duesterhoeft,A., Bancroft,I., Mewes,H.W., Rudd,S., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103638)
  AUTHORS   EU Arabidopsis sequencing project.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
COMMENT     Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1..103638
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
     misc_feature    1..8764
                     /note="overlap with BAC T22P22. Please refer to this BAC
                     for annotation and analysis"
     gene            complement(8151..8375)
                     /gene="F14F18_10"
     exon            complement(8151..8375)
                     /gene="F14F18_10"
                     /number=1
     gene            complement(join(8151..8375,8495..8551,8650..8678,
                     8890..9346))
                     /gene="F14F18_10"
     CDS             complement(join(8151..8375,8495..8551,8650..8678,
                     8890..9346))
                     /gene="F14F18_10"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB87657.1"
                     /db_xref="GI:7573351"
                     /translation="MAATVLGHFSPITSSLKPPSRHRRFLHHNLSPKTPLQSSLIKFG
                     AENSEPQPPRPLPETDCPVPEDQOPINEYQSLSTSPFFSWASGDLIEYSTRLFLTGAS
                     FAFFVGLPVSWFGSIGPEYEPVKRILAASSSGTFVVTLAVVRMYLGWAYVDEETGWYD
                     GQVWKTPEVLARDRLGSFSVSKPVLARLKNTLVLGLSLILVINLGDSPIATSYRTY
                     RDPRDRSSLPIPGAYNDETARTFEPEAFCGEPSSDLL"
     gene            complement(8376..8494)
                     /gene="F14F18_10"
                     /number=1
     intron          complement(8495..8551)
                     /gene="F14F18_10"
                     /number=2
     exon            complement(8552..8649)
```

```
ILFYPTLLYNLVRFKLQSQFRCYAVDYRNFELFTKISEYQFGSDEFSSMQKDVPRLKKL
GVGGVTTLNEPYETLVPSSLXYSAYEMEHLVITPTRDYLFAPSIVDITLAVNFIHKNALL
GKTTVYHCKAGRGRSTTVVLCYLIEHKSMTVAAAFEHVRSIRPRVLLHPSQRKVSLDR
FIDMLHQRDNRITKMKT"

CDS   complement(join(16726..16824,17130..17372,17821..18332,
      18890..19082,19231..19344))
      /note="gene_id:MIK19.7"
      pir||T13434
      similar to unknown protein"
      /codon_start=1
      /evidence=not_experimental
      /protein_id="BAB09800.1"
      /db_xref="GI:10176766"
      /translation="MKNSKCNLIDSKLEHHHLCGSKHCPGCGRMIQAATKPNWVGLP
      AGVKPDPTDQELIEHLEAKVKGKEENKKWSSSHPLIDEFIPTIDGEDGICYTHPQKLP
      GVTRDGLSKHFHFKPSRAYTTGTRKRRKIIQTDHDSELTGSSETRHKTGKTRPVMIN
      GQQRGCKKILVLYTNPGKNRPEKTNWVMHQYHLGINEEREGELVVSKIFYQTQPRQ
      CVSNINWSDHHGSKDVIGIGVGDEISSVAATLQSLGSGDVVSRVNMHPHRSFDEGTA
      EASKGRENQHVSGTCEEVHDGITTSSMGSHHMHIHDHHNQHHQIGDREEFHMSSSYPMT
      PTITSQHESIPHVTSMPFQGSVVWFGIRRPENYGLYHSYVYVRRKFLTNLCI"

CDS   join(22516..22617,22808..22854,22935..22974,23115..23324,
      23408..23566,23656..23709,23800..23998,24082..24344,
      24431..24431,24742..24933)
      /note="gene_id:MIK19.8"
      /codon_start=1
      /evidence=not_experimental
      /product="pyrophosphate-dependent
      phosphofructo-1-kinase-like protein"
      /protein_id="BAB09881.1"
      /db_xref="GI:10176767"
      /translation="MSSPRSNRPKIVNGPGYILQDVPHLIDYLPDLPTPNPLQDNP
      AYSVKQYFVHADDSVPEKVVVHKDSPRGVHFRRAGPROKYFESDEVHACIVTCGGL
      CPGLNTVIREVVSSLSYMGVGKRILGIDVSLGGYIRGFYAKNTIPLNSKVVNDIHKRGG
      TIIGTSRGGHDTNKIVDSIQDRGINQVYIGGDGTQRGASVIFEEIRRRLKVVAVGI
      PKTIONDIPVIDKSPGFDTAVEEAQRAINAAHVEARSNENGIGFVKLMGRYSGYIAMY
      ATLASRVDCCLLIPESPYLEGEGGLEFIERRLKDHGHMVTVLAEGAGQDLMCKSME
      STPMDASGNKLLKDVGLWLSQSIKDHFKKNKNMWNLKYIDPTYMIRAVPSNASDNYC
      TLLAQSAVHGAMAGYTGYTSGLVNGRQTYIPFYRIETQNNVVITDRMWARLLSSTNQ
      PSFLGPKDTSEKKELPETPLDDGAVDIPVTKEVTK"

CDS   complement(join(25739..25765,25861..25942,26048..26126,
      26210..26279,26372..26494,26552..26704,26807..26866,
      26966..27170,27390..27474,27561..27690,27813..27843))
      /note="gene_id:MIK19.9"
      pir||T06010
      similar to unknown protein"
      /codon_start=1
      /evidence=not_experimental
      /protein_id="BAB09882.1"
      /db_xref="GI:10176768"
      /translation="MNISVENPVFVHEDSTIQKTGELRLDSDIPMSKISSDDEVFLAP
      EMNAFGRQFRDYTDTNSEROKSVEHFYATOHTNOTLDFVOKMRSEYGKLDRMVNIWE
      CCELSKEVVDESDPDLDEPOIOHLLOSAEAIRKDYFNEDWLHLTALIHDLGKVLLPQ
      EGGLPOWAVVGDTFPVGCAPDESNVHIKYFMENPDDNNPKYNTKAGIYSEGCGLENVL
      MSWGHDDYMYLVAKENGSTLPSPGLFIIRYHSFYPLHKAGAYTHLMNEEDKENLKWLH
      VFNKYDLYSKSKVHHVNVEKVKPYYMSLIKKYFPENLRW"

CDS   join(29030..29383,29492..29617,29692..30006,30088..30210,
      30295..30693)
      /note="gene_id:MIK19.10"
      sp|P54969|
      /codon_start=1
      /evidence=not_experimental
      /product="IAA-amino acid hydrolase homolog 1 precursor"
      /protein_id="BAB09883.1"
      /db_xref="GI:10176769"
      /translation="MALNNELTFQLLLLLRVSSESPWIVAGDVSRKIPINFLEAKSP
      EVFDSMVRIRKIHENPELGYEEFETSKFIRSELDLIGVKYRFPVAITGIYGIGTGE
      PPFVALRADMDALPLQEAVEWEHKKKSKNPGWHHACGHDGHVAMLLGAAKILQQHROHLQ
      GTVVLFQPAEEGLSGAKMREEGALKNVEAILGEIIHLSPRIPPFGKAASLAGSFMAGAG
      AFEAVTGKGGHAAIPQHTIDPVVAASSIVLSLQHLVSREIDPSDSKVVTVKVNGGN"
```

```
Query Match           2.08;   Score 20; DB 12; Length 84129;
Best Local Similarity 100.0%;          Pred. No. 7.9; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;
```

```
     exon            /number=2
                     complement(8650..8678)
                     /gene="F14F18_10"

     intron          /number=3
                     complement(8679..8889)
                     /gene="F14F18_10"

     exon            /number=3
                     complement(8890..9346)
                     /gene="F14F18_10"

     gene            complement(join(11457..11637,11711..11760,11866..12033,
                     12133..12211,12407..12507,12606..12674,12823..12863,
                     13089..13356,13438..14056,14134..14237,14416..14856,
                     14953..15048,15122..15565))
                     /gene="F14F18_20"

     exon            /number=1
                     complement(11457..11637)
                     /gene="F14F18_20"

     CDS             complement(join(11457..11637,11711..11760,11866..12033,
                     12133..12211,12407..12507,12606..12674,12823..12863,
                     13089..13356,13438..14056,14134..14237,14416..14856,
                     14953..15048,15122..15565))
                     /gene="F14F18_20"
                     /codon_start=1
                     /product="MAP3K delta-1 protein kinase"
                     /note="The mRNA covering this gene is only partial. This
                     has allowed the definition of the carboxy terminus of the
                     protein. The amino terminus remains predicted but
                     prediction is supported by different gene models."
                     /protein_id="CAB87658.1"
                     /db_xref="GI:7573352"
                     /translation="MSKMKHLLRKLHTGGSSCVGGGFPADHHRLDDDSTRPMIDPSPILS
                     TSPSPASTSSVSSSSGFGNASTTMPRLDTEFPVGRDLTPAVDGVDFNLMEEEYQVQLAMA
                     ISVSDPPRENADTAQLDAAKRISLGVSAPVTDADSAVDFLSLRYWSGCSISEGLKEL
                     GSWLLPAVFSSLRVFLNPNCGHKVINYDQKVRDGFYDVYGITSNSLSQGKMPLLVDLQ
                     AISIDNDVEVILVNRLIIDPELQELERRVFALASECPDFAPGQVSSDLTQKIANIVV
                     EQMGGPVENADEALRRWMLRSYELRNSLNTTILPLGRVNVGLARHRALLFKVLADRIN
                     LPCMLVKGSYYGTDGDCAVNLIKLDDKSPNSFFTCSEYIIDLMGPAGALIPSEVPSSF
                     LPVSCTDTRVPENDLSLQHSSPVLEKEIETPAFSVSKEADSRSGMVANFFTGNQEEN
                     SDRCAVEKHQTERFEHDFGKLMHSQOISGENMPPFSGKPTCAQKVKVKNVSKYVISAA
                     KNPEFAQKLHAVLLESGASPPDLFMDINPHNLRGKNLLQELRQESSNSMVSGIPCYP
                     EKVAEQLRESERNPFAESYOQSVEVDLSMKRNFDLDNTGKASSSENMEVGTADGESAV
                     CSHDQGINPLLGEAAKWEIMWEDLQIGERICIGSVQEVYRAEWNGTEVAAVKFLDQD
                     FSGDALTQFKSEIEIMLRLRHPNVVLPFMGAVTRPPNFSILTEFLPRGSLYRLLHRPNH
                     OLDEKRMRMALDVAKGMNYLHTHSHPTVVHRDLKSPNLLVDKNWVKVCVCDFGLSRMKH
                     HTYLSKSTAGTPEWMAPEVLRNEPANEKCDVYSFGVILWELATSRVPWKGLNPMQVV
                     GAVFGQNRLEIPDDIDLTVAQIIRECWQT"

     gene            11457..15565
                     /gene="F14F18_20"

     exon            /number=1
                     complement(11457..11637)
                     /gene="F14F18_20"

     intron          /number=1
                     complement(11638..11710)
                     /gene="F14F18_20"

     exon            /number=2
                     complement(11711..11760)
                     /gene="F14F18_20"

     intron          /number=2
                     complement(11761..11865)
                     /gene="F14F18_20"

     exon            /number=3
                     complement(11866..12033)
                     /gene="F14F18_20"

     intron          /number=3
                     complement(12034..12132)
                     /gene="F14F18_20"

     exon            /number=4
                     complement(12133..12211)
                     /gene="F14F18_20"

     intron          /number=4
                     complement(12212..12406)
                     /gene="F14F18_20"

     exon            /number=5
                     complement(12407..12507)
                     /gene="F14F18_20"

     intron          /number=5
                     complement(12508..12605)
                     /gene="F14F18_20"

     exon            /number=6
                     complement(12606..12674)
                     /gene="F14F18_20"

     intron          /number=6
                     complement(12675..12822)
                     /gene="F14F18_20"

     exon            /number=7
                     complement(12823..12863)
                     /gene="F14F18_20"

     intron          /number=7
                     complement(12864..13088)
                     /gene="F14F18_20"

     exon            /number=8
                     complement(13089..13356)
                     /gene="F14F18_20"

     intron          /number=8
                     complement(13357..13437)
                     /gene="F14F18_20"

     exon            /number=9
                     complement(13438..14056)
                     /gene="F14F18_20"

     intron          /number=9
                     complement(14057..14133)
                     /gene="F14F18_20"

     exon            /number=10
                     complement(14134..14237)
                     /gene="F14F18_20"

     intron          /number=10
                     complement(14238..14415)
                     /gene="F14F18_20"

     exon            /number=11
                     complement(14416..14856)
                     /gene="F14F18_20"

     intron          /number=11
                     complement(14857..14952)
                     /gene="F14F18_20"

     exon            /number=12
                     complement(14953..15048)
                     /gene="F14F18_20"

     intron          /number=12
                     complement(15049..15121)
                     /gene="F14F18_20"

     exon            /number=13
                     complement(15122..15565)
                     /gene="F14F18_20"

     gene            16565..17936
                     /gene="F14F18_30"

     CDS             complement(join(16565..16786,17110..17447,17579..17936))
                     /gene="F14F18_30"
                     /codon_start=1
                     /product="putative protein"
                     /note="similarity to NLI interacting factor isoform T1,
                     Gallus gallus, EMBL:AF189773"
                     /protein_id="CAB87659.1"
                     /db_xref="GI:7573353"
                     /translation="MEDDSPSSRRDLDAQNPYDRLLALDTSTVDPNCNLDSVSAIYLA
                     MKSSKLECVDERGQDSLIISVCMEDEDEDEELDEDFDPYLFIKNLPNLSSVVPTFRPVLL
                     PKQTRSCPPISLVLDLDETLVHSTLEPCGEVDFTPVNFNEEEHMVVRCRPHLKEFM
                     ERVSRLFEIIFTASOSIVAEQLLNVLDPKRKLFRHRVYRDSCVFFDGNYLKDLSVLG
                     RDLSRVIVDNSPQAFGFQVENGVPIESWFNDPSDKELLHLIPLFESLIGVEDVRPMI
                     AKKFNLREKIDAAVAAPEYPAEAGDPFER"

     exon            /number=1
                     complement(16565..16786)
                     /gene="F14F18_30"

     intron          /number=1
                     complement(16787..17109)
                     /gene="F14F18_30"

     exon            /number=2
                     complement(17110..17447)
                     /gene="F14F18_30"

     intron          /number=2
                     complement(17448..17578)
                     /gene="F14F18_30"
```

RESULT 22

```
Db   90372  ATTTGGAAGAGTTTCAGACC  90391
Qy     760  atttggaagagtttcaagacc  779
            |||||||||||||||||||||
```

Query Match   2.0%;   Score   20;   DB 13;   Length 103638;
Best Local Similarity 100.0%;   Pred. No. 7.9;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
LOCUS       HS173D1/c      117338 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from clone 173D1 on chromosome 1p36.21-36.33.
            Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL031984
VERSION     AL031984.13  GI:4678432
KEYWORDS    HTG; CpG Island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117338)
  AUTHORS   Moore,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 24, 1999 this sequence version replaced gi:4584747.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source database:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL.
            This sequence is the entire insert of clone 173D1. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            173D1 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..117338
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.21-36.33"
                     /clone="RP1-173D1"
                     /clone_lib="RPCI-1"
     repeat_region   91..142
                     /note="13 copies 4 mer aggg 75% conserved"
     repeat_region   585..992
                     /note="2 copies 54 mer 96% conserved"
     repeat_region   1728..1903
                     /note="17 copies 24 mer 97% conserved"
     repeat_region   1909..2078
                     /note="MER45 repeat: matches 1..178 of consensus"
     repeat_region   4058..4357
                     /note="MIR repeat: matches 28..187 of consensus"
     repeat_region   4464..4571
                     /note="AluY repeat: matches 1..298 of consensus"
     repeat_region   4726..4844
                     /note="MIR repeat: matches 21..151 of consensus"
     repeat_region   5494..5786
                     /note="AluSg repeat: matches 1..292 of consensus"
     repeat_region   6503..6807
                     /note="AluSg repeat: matches 1..311 of consensus"
     repeat_region   6809..6857
                     /note="AluJo repeat: matches 1..150 of consensus"
     repeat_region   6858..7181
                     /note="MIR repeat: matches 101..311 of consensus"
     repeat_region   7182..7299
                     /note="AluSx repeat: matches 1..262 of consensus"
     repeat_region   8932..8993
                     /note="MIR repeat: matches 150..262 of consensus"

     repeat_region   8994..9305
                     /note="MER53 repeat: matches 126..189 of consensus"
     repeat_region   9306..9440
                     /note="AluSq repeat: matches 1..311 of consensus"
     repeat_region   10094..10654
                     /note="MER53 repeat: matches 1..126 of consensus"
     repeat_region   13515..13745
                     /note="L2 repeat: matches 2152..2746 of consensus"
     repeat_region   13746..14051
                     /note="L2 repeat: matches 1886..2084 of consensus"
     repeat_region   14052..14530
                     /note="AluSx repeat: matches 1..306 of consensus"
     repeat_region   14561..14871
                     /note="L2 repeat: matches 2084..2680 of consensus"
     repeat_region   16630..16703
                     /note="AluSx repeat: matches 1..311 of consensus"
     misc_feature    complement(16703..17036..17418)
                     /note="MIR repeat: matches 101..176 of consensus"
     misc_feature    19534..19676
                     /note="match: STS G11038"
     repeat_region   19811..20021
                     /note="MIR repeat: matches 26..168 of consensus"
     repeat_region   20103..20402
                     /note="L2 repeat: matches 2521..2705 of consensus"
     repeat_region   20551..20620
                     /note="AluSx repeat: matches 1..300 of consensus"
     repeat_region   20681..20878
                     /note="L2 repeat: matches 2632..2701 of consensus"
     repeat_region   22445..22496
                     /note="MIR repeat: matches 28..238 of consensus"
     repeat_region   22496..22714
                     /note="12 copies 2 mer ca 100% conserved"
     repeat_region   22969..23143
                     /note="MIR repeat: matches 2..223 of consensus"
     repeat_region   23333..23360
                     /note="MIR repeat: matches 71..232 of consensus"
     repeat_region   24447..24757
                     /note="2 copies 64 mer 98% conserved"
     repeat_region   25099..25400
                     /note="AluSg repeat: matches 1..309 of consensus"
     repeat_region   25523..25637
                     /note="AluSx repeat: matches 1..303 of consensus"
     repeat_region   27898..27939
                     /note="Charlie5 repeat: matches 5..121 of consensus"
     repeat_region   29266..29305
                     /note="7 copies 6 mer tcctcc 95% conserved"
     repeat_region   25767..26056
                     /note="AluSx repeat: matches 1..291 of consensus"
     repeat_region   26059..26270
                     /note="10 copies 4 mer aagg 80% conserved"
     misc_feature    31618..31978
                     /note="match: EST AI350672"
     repeat_region   33293..33452
                     /note="MIR repeat: matches 48..203 of consensus"
     repeat_region   33832..34011
                     /note="AluJb repeat: matches 1..312 of consensus"
     repeat_region   33832..33999
                     /note="90 copies 2 mer ac 71% conserved"
     repeat_region   33832..34023
                     /note="7 copies 24 mer 73% conserved"
     repeat_region   33832..33993
                     /note="3 copies 64 mer 72% conserved"
     repeat_region   33832..34503
                     /note="3 copies 54 mer 74% conserved"
     repeat_region   34018..34497
                     /note="9 copies 54 mer 67% conserved"
     repeat_region   34018..34505
                     /note="20 copies 24 mer 66% conserved"
     repeat_region   34018..34491
                     /note="122 copies 4 mer caca 65% conserved"
```

```
RESULT  23
LOCUS       AC027795   148508 bp    DNA              HTG       03-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-235B15 map 18, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC027795
VERSION     AC027795.2  GI:7684508
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
```

```
Qy       10 caagattcaaacccagctca 29
            ||||||||||||||||||||
Db    68061 CAAGATTCAAACCCAGCTCA 68042

Query Match          2.0%;  Score 20;  DB 92;  Length 117338;
Best Local Similarity 100.0%;  Pred. No. 7.9;
Matches  20:  Conservative  0:  Mismatches  0:  Indels  0:  Gaps  0:
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148508)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-235B15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148508)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,D., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zalnoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 3, 2000 this sequence version replaced gi:7382637.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----------------- Project Information
            Center project name: L9034
            Center clone name: 235_B15
            ----------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 13822 bases at least Q40
            Consensus quality: 143906 bases at least Q30
            Consensus quality: 145597 bases at least Q20
            Insert size: 157000; agarose-fp
            Insert size: 146908; sum-of-contigs
            Quality coverage: 4.3 in Q20 bases; agarose-fp
            Quality coverage: 4.6 in Q20 bases; sum-of-contigs
            -----------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *               1   1262: contig of 1262 bp in length
            *            1263   1362: gap of    100 bp
            *            1363   2466: contig of 1104 bp in length
            *            2467   2566: gap of    100 bp
            *            2567   3867: contig of 1301 bp in length
            *            3868   3967: gap of    100 bp
            *            3968   4372: contig of  405 bp in length
            *            4373   4472: gap of    100 bp
```

```
     repeat_region   34025..34504
                     /note="237 copies 2 mer ca 66% conserved"
     repeat_region   34028..34507
                     /note="48 copies 10 mer acacacacac 65% conserved"
     repeat_region   34065..34512
                     /note="80 copies 6 mer cacaca 65% conserved"
     repeat_region   34501..34914
                     /note="7 copies 64 mer 70% conserved"
     repeat_region   34501..34914
                     /note="3 copies 138 mer 70% conserved"
     repeat_region   34508..34837
                     /note="33 copies 10 mer acacacac 66% conserved"
     repeat_region   34510..34839
                     /note="55 copies 6 mer acacac 66% conserved"
     repeat_region   34511..34846
                     /note="14 copies 24 mer 66% conserved"
     repeat_region   34519..34842
                     /note="6 copies 54 mer 67% conserved"
     repeat_region   34519..34838
                     /note="160 copies 2 mer ca 66% conserved"
     repeat_region   35512..35541
                     /note="15 copies 2 mer tt 90% conserved"
     misc_feature    complement(36540..36839)
                     /note="match: EST AA295567"
     repeat_region   36962..37180
                     /note="MIR repeat: matches 9. .251 of consensus"
     repeat_region   37142..37185
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   37989..38163
                     /note="L2 repeat: matches 2655. .2698 of consensus"
     repeat_region   39487..39536
                     /note="MIR repeat: matches 47. .236 of consensus"
     repeat_region   39487..39536
                     /note="5 copies 10 mer gtgtgtgt 90% conserved"
     repeat_region   39487..39528
                     /note="21 copies 2 mer gt 98% conserved"
     repeat_region   39487..39534
                     /note="12 copies 4 mer gtgt 92% conserved"
     repeat_region   39831..40133
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   40095..41105
                     /note="Charlie5 repeat: matches 1. .62 of consensus"
     repeat_region   40574..40710
                     /note="MER3 repeat: matches 9. .149 of consensus"
     repeat_region   40684..40748
                     /note="MIR repeat: matches 32. .222 of consensus"
     misc_feature    complement(41047..41397)
                     /note="matching STS L31008"
     repeat_region   41244..41387
                     /note="match: STS L31008"
     repeat_region   41263..41538
                     /note="72 copies 2 mer tc 63% conserved"
     repeat_region   41264..41527
                     /note="46 copies 6 mer ttcttt 57% conserved"
     repeat_region   41271..41465
                     /note="11 copies 24 mer 61% conserved"
     repeat_region   
                     /note="39 copies 39 mer 73% conserved"
```

```
FEATURES             Location/Qualifiers
     source          1. 148508
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-235B15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1262
                     /note="assembly_fragment"
     misc_feature    1363. .2466
                     /note="assembly_fragment"
     misc_feature    2567. .3867
                     /note="assembly_fragment"
     misc_feature    3968. .4372
                     /note="assembly_fragment"
     misc_feature    4473. .7936
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    4473. .7936
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    12074. .16804
                     /note="assembly_fragment"
     misc_feature    8037. .11973
                     /note="assembly_fragment"
     misc_feature    22644. .27921
                     /note="assembly_fragment"
     misc_feature    16905. .22543
                     /note="assembly_fragment"
     misc_feature    28022. .33883
                     /note="assembly_fragment"
     misc_feature    33984. .39792
                     /note="assembly_fragment"
     misc_feature    49549. .59577
                     /note="assembly_fragment"
     misc_feature    39993. .49448
                     /note="assembly_fragment"
     misc_feature    59678. .69844
                     /note="assembly_fragment"
     misc_feature    69945. .95609
                     /note="assembly_fragment"
     misc_feature    95710. .120155
                     /note="assembly_fragment"
     misc_feature    120256. .148508
                     /note="assembly_fragment"

*   7473   7936: contig of 3464 bp in length
*   7937   8036: gap of 100 bp
*   8037  11973: contig of 3937 bp in length
*  11974  12073: gap of 100 bp
*  12074  16804: contig of 4731 bp in length
*  16805  16904: gap of 100 bp
*  16905  22543: contig of 5639 bp in length
*  22544  22643: gap of 100 bp
*  22644  27921: contig of 5278 bp in length
*  27922  28021: gap of 100 bp
*  28022  33883: contig of 5862 bp in length
*  33884  33983: gap of 100 bp
*  33984  39792: contig of 5809 bp in length
*  39793  39892: gap of 100 bp
*  39893  49448: contig of 9556 bp in length
*  49449  49548: gap of 100 bp
*  49549  59577: contig of 10029 bp in length
*  59578  59677: gap of 100 bp
*  59678  69844: contig of 10167 bp in length
*  69845  69944: gap of 100 bp
*  69945  95609: contig of 25665 bp in length
*  95610  95709: gap of 100 bp
*  95710 120155: contig of 24446 bp in length
* 120156 120255: gap of 100 bp
* 120256 148508: contig of 28253 bp in length.

BASE COUNT    44364 a   29387 c   30187 g   42568 t   1602 others
ORIGIN
```

```
Query Match                          2.0%;  Score 20;  DB 71;  Length 148508;
Best Local Similarity   100.0%;  Pred. No. 7.9;
Matches  20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    304 agaaatcatcttccttggaa 323
          ||||||||||||||||||||
Db   9577 AGAAATCATCTTCCTTGGAA 9558

RESULT  24
AC015525
LOCUS       AC015525    149180 bp    DNA             HTG      12-MAR-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-110C15 map 15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC015525
VERSION     AC015525.3  GI:7108019
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149180)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 15, clone RP11-110C15
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 149180)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,N., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,K., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6563530.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
            Center project name: L4276
            Center clone name: 110_C_15
            -------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 132714 bases at least Q40
            Consensus quality: 142719 bases at least Q30
            Consensus quality: 146304 bases at least Q20
            Insert size: 156000; agarose-fp
            Insert size: 148380; sum-of-contigs
            Quality coverage: 5.2 in Q20 bases; agarose-fp
            Quality coverage: 5.5 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
RESULT  25
AC069420/c

Db 135751 GACCAGGTGGACAGGGTCCT 135770
                    ||||||||||||||||||||
Qy      449 gaccaggtggacaggtcct 468

Matches  20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
Best Local Similarity  100.0%;  Pred. No. 7.9;
Query Match              2.0%;  Score 20;  DB 63;  Length 149180;
```

```
ORIGIN
BASE COUNT     42836 a   32086 c   31253 g   42201 t      804 others
            /note="assembly_fragment"
misc_feature   102434. .149180
            /note="assembly_fragment"
misc_feature   59779. .102333
            clone_end:SP6
            vector_side:right"
            clone_end:T7
            /note="assembly_fragment"
misc_feature   40398. .59678
            vector_side:right"
            /note="assembly_fragment"
misc_feature   3740. .40297
            /note="assembly_fragment"
misc_feature   22412. .30639
            /note="assembly_fragment"
misc_feature   14972. .22311
            /note="assembly_fragment"
misc_feature   8538. .14871
            /note="assembly_fragment"
misc_feature   2489. .8437
            /note="assembly_fragment"
misc_feature   1. .2388
            /clone_lib="RPCI-11 Human Male BAC"
            /clone="RPII-11OC15"
            /map="15"
            /chromosome="15"
            /db_xref="taxon:9606"
            /organism="Homo sapiens"
source         1. 149180
            Location/Qualifiers
FEATURES
```

```
            102434  149180: contig of 46747 bp in length.
            102433  102433: gap of 100 bp in length
            59779  102433: contig of 42555 bp in length
            59778: 59778: gap of 100 bp in length
            40398  59677: contig of 19281 bp in length
            40298  40397: gap of 100 bp in length
            30740  40297: contig of 9558 bp in length
            30640  30739: gap of 100 bp in length
            22412  30639: contig of 8228 bp in length
            22312  22411: gap of 100 bp in length
            14972  22311: contig of 7340 bp in length
            14872  14971: gap of 100 bp in length
            8538   14871: contig of 6334 bp in length
            8438   8537: gap of 100 bp in length
            2489   8437: contig of 5949 bp in length
            2389   2488: gap of 100 bp in length
            1      2388: contig of 2388 bp in length
        * be preserved.
        * as soon as it is available and the accession number will
        * This record will be updated with the finished sequence
        * runs of N, but the exact sizes of the gaps are unknown.
```

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 150683)
AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
           Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
           Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
           Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
           Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
           Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.J., Chen,G., Chen,R.,
           Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
           Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
           Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
           Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
           Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
           Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
           Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
           Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
           Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
           Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,J.,
           Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
           Jackson,L.E., Jacobson,D., Jia,Y., Johnson,K., Jolivet,S.,
           Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
           Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
           Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
           Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
           Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
           Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
           Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
           Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
           Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
           Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
           Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
           Puazo,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,S.,
           Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
           Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
           Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K.,
           Tang,H., Tansey,J., Taylor,C., Taylor,P., Telfrod,B., Thomas,N.,
           Thomas,S., Usmani,K., Vasquez,L., Veta,V., Villalon,D., Vinson,R.,
           Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
           Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
           Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
           and Gibbs,R.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 150683)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (30-MAY-2000) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Sep 10, 2000 this sequence version replaced gi:9795456.
           ------------ Genome Center
           Center: Baylor College of Medicine
           Center code: BCM
           Web site: http://www.hgsc.bcm.tmc.edu/
           Contact: hgsc-help@bcm.tmc.edu
           ------------ Project Information
           Center project name: HGGN
           Center clone name: RPI1-683J13
           ------------ Summary Statistics
           Assembly program: Phrap; version 0.990329
           Consensus quality: 125821 bases at least Q40
           Consensus quality: 138453 bases at least Q30
           Consensus quality: 144027 bases at least Q20
           Estimated insert size: 142714; sum-of-contigs estimation
           Quality coverage: 0x in Q20 bases; agarose-fp estimation
           Quality coverage: 4.1x in Q20 bases; sum-of-contigs estimation
           --------------
           * NOTE: Estimated insert size may differ from sequence length
           *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 23 contigs. The true order of the pieces
```

```
ACCESSION  AC069420
VERSION    AC069420.6  GI:11047548
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.

ACCESSION  AC069420
LOCUS      AC069420   150683 bp   DNA                HTG   11-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-683J13, WORKING DRAFT
           SEQUENCE, 23 unordered pieces.
```

LOCUS   AC018876   165493 bp   DNA   HTG   07-JUL-2000

RESULT 26
AC018876/c

Qy   449 gaccaggtggacaggtcct 468
          |||||||||||||||||||
Db   878 GACCAGGTGGACAGGTCCT 859

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Best Local Similarity   100.0%; Pred. No. 7.9;
Query Match   2.0%;   Score 20;   DB 74;   Length 150683;

ORIGIN
BASE COUNT   40413 a   33700 c   32450 g   41878 t   2242 others

FEATURES   Location/Qualifiers
     source   1..150683
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-683J13"

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1   24905: contig of 24905 bp in length
*   24906   25005: gap of unknown length
*   25006   38744: contig of 13739 bp in length
*   38745   38844: gap of unknown length
*   38845   53622: contig of 14778 bp in length
*   53623   53722: gap of unknown length
*   53723   65642: contig of 11920 bp in length
*   65643   65742: gap of unknown length
*   65743   73486: contig of 7744 bp in length
*   73487   73586: gap of unknown length
*   73587   81893: contig of 8307 bp in length
*   81894   81993: gap of unknown length
*   81994   88696: contig of 6703 bp in length
*   88697   88796: gap of unknown length
*   88797   97825: contig of 9029 bp in length
*   97826   97925: gap of unknown length
*   97926   103924: contig of 5999 bp in length
*   103925   104024: gap of unknown length
*   104025   109912: contig of 5888 bp in length
*   109913   110012: gap of unknown length
*   110013   116598: contig of 6586 bp in length
*   116599   116698: gap of unknown length
*   116699   120672: contig of 3974 bp in length
*   120673   120772: gap of unknown length
*   120773   125139: contig of 4367 bp in length
*   125140   125229: gap of unknown length
*   125229   128431: contig of 3092 bp in length
*   128432   128531: gap of unknown length
*   128532   131187: contig of 2756 bp in length
*   131188   131287: gap of unknown length
*   131288   134177: contig of 2890 bp in length
*   134178   134277: gap of unknown length
*   134278   137482: contig of 3205 bp in length
*   137483   137582: gap of unknown length
*   137583   138864: contig of 1282 bp in length
*   138865   138964: gap of unknown length
*   138965   142546: contig of 3582 bp in length
*   142547   142646: gap of unknown length
*   142647   145006: contig of 2360 bp in length
*   145007   145106: gap of unknown length
*   145107   147639: contig of 2533 bp in length
*   147640   147739: gap of unknown length
*   147740   149420: contig of 1681 bp in length
*   149421   149520: gap of unknown length
*   149521   150683: contig of 1163 bp in length.

DEFINITION Homo sapiens chromosome 18 clone RP11-289E15, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC018876
VERSION   AC018876.4 GI:7230959
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 165493)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 165493)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Mar 13, 2000 this sequence version replaced gi:6850795.
            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Center project name: H_NH0289E15
            ---------------- Project Information ----------------
            ---------------- Summary Statistics ----------------
            Sequencing vector: M13; 81%
            Sequencing vector: plasmid; 19%
            Chemistry: Dye-primer ET; 81% of reads
            Chemistry: Dye-terminator Big Dye; 19% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157046 bases at least Q40
            Consensus quality: 160223 bases at least Q30
            Consensus quality: 161942 bases at least Q20
            Insert size: 174000; agarose-fp
            Insert size: 164793; sum-of-contigs
            Quality coverage: 4.36 in Q20 bases; agarose-fp
            Quality coverage: 4.61 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1322: contig of 1322 bp in length
            *   1323   1422: gap of unknown length
            *   1423   8681: contig of 7259 bp in length
            *   8682   8781: gap of unknown length
            *   8782   28642: contig of 19861 bp in length
            *   28643   28742: gap of unknown length
            *   28743   46484: contig of 17742 bp in length
            *   46485   46584: gap of unknown length
            *   46585   68690: contig of 22106 bp in length
            *   68691   68790: gap of unknown length
            *   68791   99194: contig of 30404 bp in length
            *   99195   99294: gap of unknown length
            *   99295   131431: contig of 32137 bp in length
            *   131432   131531: gap of unknown length
            *   131532   165493: contig of 33962 bp in length.
FEATURES   Location/Qualifiers
     source   1..165493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-289E15"
ORIGIN
BASE COUNT   52008 a   30658 c   30763 g   51358 t   706 others

```
Qy    578 gcttcaccctttgctggtgg 597
          |||||||||||||||||||||
Db  11874 GCTTCACCCTTTGCTGGTGG 11855
```

```
Query Match           2.0%;   Score 20;  DB 65;  Length 165493;
Best Local Similarity  100.0%;   Pred. No. 7.9;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
RESULT 27
LOCUS       AC021927   168997 bp    DNA            HTG        03-MAR-2000
DEFINITION  Homo sapiens clone RP11-29P22, WORKING DRAFT SEQUENCE, 26 unordered
            pieces.
ACCESSION   AC021927
VERSION     AC021927.3  GI:7144973
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168997)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-29P22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168997)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,P., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6984426.
            All repeats were identified using RepeatMasker!
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------------------------------------------------
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ------------------------------------------------------------
            Center project name: L4856
            Center clone name: 29_P_22
            ------------------------------------------------------------
            Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 156612 bases at least Q40
            Consensus quality: 162417 bases at least Q30
            Consensus quality: 164820 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 166497; sum-of-contigs
            Quality coverage: 3.6 in Q20 bases; agarose-fp
            Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            ------------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     1046: contig of 1046 bp in length
            *   1047     1146: gap of   100 bp
            *   1147     2513: contig of 1367 bp in length
            *   2514     2613: gap of   100 bp
            *   2614     3854: contig of 1241 bp in length
            *   3855     3954: gap of   100 bp
            *   3955     5409: contig of 1455 bp in length
            *   5410     5509: gap of   100 bp
            *   5510     8344: contig of 2835 bp in length
            *   8345     8444: gap of   100 bp
            *   8445    11304: contig of 2860 bp in length
            *  11305    11404: gap of   100 bp
            *  11405    14151: contig of 2747 bp in length
            *  14152    14251: gap of   100 bp
            *  14252    17174: contig of 2923 bp in length
            *  17175    17274: gap of   100 bp
            *  17275    19939: contig of 2665 bp in length
            *  19940    20039: gap of   100 bp
            *  20040    23411: contig of 3372 bp in length
            *  23412    23511: gap of   100 bp
            *  23512    27130: contig of 3619 bp in length
            *  27131    27230: gap of   100 bp
            *  27231    29770: contig of 2540 bp in length
            *  29771    29870: gap of   100 bp
            *  29871    36121: contig of 6251 bp in length
            *  36122    36221: gap of   100 bp
            *  36222    41978: contig of 5757 bp in length
            *  41979    42078: gap of   100 bp
            *  42079    49159: contig of 7081 bp in length
            *  49160    49259: gap of   100 bp
            *  49260    57801: contig of 8542 bp in length
            *  57802    57901: gap of   100 bp
            *  57902    64360: contig of 6459 bp in length
            *  64361    64460: gap of   100 bp
            *  64461    72524: contig of 8064 bp in length
            *  72525    72624: gap of   100 bp
            *  72625    81728: contig of 9104 bp in length
            *  81729    81828: gap of   100 bp
            *  81829    92259: contig of 10431 bp in length
            *  92260    92359: gap of   100 bp
            *  92360   100902: contig of 8543 bp in length
            *  100903  101002: gap of   100 bp
            *  101003  110082: contig of 9080 bp in length
            *  110083  110182: gap of   100 bp
            *  110183  124956: contig of 14774 bp in length
            *  124957  125056: gap of   100 bp
            *  125057  139552: contig of 14496 bp in length
            *  139553  139652: gap of   100 bp
            *  139653  153066: contig of 13413 bp in length
            *  153066  153165: gap of   100 bp
            *  153166  168997: contig of 15832 bp in length.
FEATURES             Location/Qualifiers
     source          1..168997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-29P22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1046
                     /note="assembly_fragment"
     misc_feature    1147..2513
                     /note="assembly_fragment"
     misc_feature    2614..3854
                     /note="assembly_fragment"
     misc_feature    3955..5409
                     /note="assembly_fragment"
     misc_feature    5510..8344
                     /note="assembly_fragment"
```

```
misc_feature   8445..11304
               /note="assembly_fragment"
misc_feature   11405..14151
               /note="assembly_fragment"
misc_feature   14252..17174
               /note="assembly_fragment"
misc_feature   17275..19939
               /note="assembly_fragment"
misc_feature   20040..23411
               /note="assembly_fragment"
misc_feature   23512..27130
               /note="assembly_fragment"
misc_feature   27231..29770
               /note="assembly_fragment"
               /note="clone_end:T7"
               /note="vector_side:left"
misc_feature   29871..36121
               /note="assembly_fragment"
misc_feature   36222..41978
               /note="assembly_fragment"
misc_feature   42079..49159
               /note="assembly_fragment"
misc_feature   49260..57801
               /note="assembly_fragment"
misc_feature   57902..64360
               /note="assembly_fragment"
misc_feature   64461..72524
               /note="assembly_fragment"
misc_feature   72625..81728
               /note="assembly_fragment"
misc_feature   81829..92259
               /note="assembly_fragment"
misc_feature   92360..100902
               /note="assembly_fragment"
misc_feature   101003..110082
               /note="assembly_fragment"
               /note="vector_side:right"
               /note="clone_end:SP6"
misc_feature   110183..124956
               /note="assembly_fragment"
misc_feature   125057..139552
               /note="assembly_fragment"
misc_feature   139653..153065
               /note="assembly_fragment"
misc_feature   153166..168997
               /note="assembly_fragment"
BASE COUNT     48830 a   34497 c   33807 g   49359 t   2504 others
ORIGIN
```

```
TITLE       Homo sapiens chromosome 4, clone RP11-285A15
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 177868)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 3, 2000 this sequence version replaced gi:7188869.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information --------------
            Center project name: L5067
            Center clone name: 285_A_15
            -------------- Summary Statistics --------------
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Assembly quality: 150228 bases at least Q40
            Consensus quality: 163346 bases at least Q30
            Consensus quality: 169409 bases at least Q20
            Insert size: 189000; agarose-fp
            Insert size: 173668; sum-of-contigs
            Quality coverage: 2.9 in Q20 bases; agarose-fp
            Quality coverage: 3.2 in Q20 bases; sum-of-contigs
            --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 43 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
            --------------
          *          1      1215: contig of 1215 bp in length
          *       1216      1315: gap of 100 bp
          *       1316      2523: contig of 1208 bp in length
          *       2524      2623: gap of 100 bp
          *       2624      3811: contig of 1188 bp in length
          *       3812      3911: gap of 100 bp
          *       3912      5309: contig of 1398 bp in length
          *       5310      5409: gap of 100 bp
          *       5410      7250: contig of 1841 bp in length
          *       7251      7350: gap of 100 bp
          *       7351      8970: contig of 1620 bp in length
          *       8971      9070: gap of 100 bp
          *       9071     10289: contig of 1219 bp in length
          *      10290     10389: gap of
```

```
RESULT 28
LOCUS       AC025191   177868 bp   DNA            HTG       03-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-285A15 map 4, WORKING DRAFT
            SEQUENCE, 43 unordered pieces.
ACCESSION   AC025191
VERSION     AC025191.2  GI:7684444
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177868)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
Qy      304 agaaatcatcttccttggaa 323
            ||||||||||||||||||||
Db   110794 AGAAATCATCTTCCTTGGAA 110813

Query Match        2.0%;   Score 20;  DB 66;  Length 168997;
Best Local Similarity 100.0%;   Pred. No. 7.9;
Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
FEATURES             Location/Qualifiers
     source          1..177868
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-285A15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1215
                     /note="assembly_fragment"
     misc_feature    1316..2523
                     /note="assembly_fragment"
     misc_feature    2624..3811
                     /note="assembly_fragment"
     misc_feature    3912..5309
                     /note="assembly_fragment"
     misc_feature    5410..7250
                     /note="assembly_fragment"
     misc_feature    7351..8970
                     /note="assembly_fragment"
     misc_feature    9071..10289
                     /note="assembly_fragment"
     misc_feature    10390..11577
                     /note="assembly_fragment"
     misc_feature    11678..12960
                     /note="assembly_fragment"
     misc_feature    13061..14891
                     /note="assembly_fragment"
     misc_feature    14992..16528
                     /note="assembly_fragment"
     misc_feature    16629..18711
                     /note="assembly_fragment"
     misc_feature    18812..20214
                     /note="assembly_fragment"
     misc_feature    20315..22282
                     /note="assembly_fragment"
     misc_feature    22383..24680
                     /note="assembly_fragment"
     misc_feature    24781..27243
                     /note="assembly_fragment"
     misc_feature    27344..29231
                     /note="assembly_fragment"
     misc_feature    29332..32345
                     /note="assembly_fragment"
```

```
  *   10390  11577: contig of 1188 bp in length
  *   11578  11677: gap of 100 bp
  *   11678  12960: contig of 1283 bp in length
  *   12961  13060: gap of 100 bp
  *   13061  14891: contig of 1831 bp in length
  *   14892  14991: gap of 100 bp
  *   14992  16528: contig of 1537 bp in length
  *   16529  16628: gap of 100 bp
  *   16629  18711: contig of 2083 bp in length
  *   18712  18811: gap of 100 bp
  *   18812  20214: contig of 1403 bp in length
  *   20215  20314: gap of 100 bp
  *   20315  22282: contig of 1968 bp in length
  *   22283  22382: gap of 100 bp
  *   22383  24680: contig of 2298 bp in length
  *   24681  24780: gap of 100 bp
  *   24781  27243: contig of 2463 bp in length
  *   27244  27343: gap of 100 bp
  *   27344  29231: contig of 1888 bp in length
  *   29232  29331: gap of 100 bp
  *   29332  32345: contig of 3014 bp in length
  *   32346  32445: gap of 100 bp
  *   32446  35005: contig of 2560 bp in length
  *   35006  35105: gap of 100 bp
  *   35106  37773: contig of 2668 bp in length
  *   37774  37873: gap of 100 bp
  *   37874  40633: contig of 2760 bp in length
  *   40634  40733: gap of 100 bp
  *   40734  43133: contig of 2400 bp in length
  *   43134  43233: gap of 100 bp
  *   43234  45935: contig of 2702 bp in length
  *   45936  46035: gap of 100 bp
  *   46036  48895: contig of 2860 bp in length
  *   48896  48995: gap of 100 bp
  *   48996  51307: contig of 2312 bp in length
  *   51308  51407: gap of 100 bp
  *   51408  55313: contig of 3906 bp in length
  *   55314  55413: gap of 100 bp
  *   55414  58208: contig of 2795 bp in length
  *   58209  58308: gap of 100 bp
  *   58309  62616: contig of 4308 bp in length
  *   62617  62716: gap of 100 bp
  *   62717  67413: contig of 4697 bp in length
  *   67414  67513: gap of 100 bp
  *   67514  70733: contig of 3220 bp in length
  *   70734  70833: gap of 100 bp
  *   70834  75405: contig of 4572 bp in length
  *   75406  75505: gap of 100 bp
  *   75506  80523: contig of 5018 bp in length
  *   80524  80623: gap of 100 bp
  *   80624  86066: contig of 5443 bp in length
  *   86067  86166: gap of 100 bp
  *   86167  92000: contig of 5834 bp in length
  *   92001  92100: gap of 100 bp
  *   92101  96663: contig of 4563 bp in length
  *   96664  96763: gap of 100 bp
  *   96764  102888: contig of 6125 bp in length
  *  102889  102988: gap of 100 bp
  *  102989  110032: contig of 7044 bp in length
  *  110033  110132: gap of 100 bp
  *  110133  117903: contig of 7771 bp in length
  *  117904  118003: gap of 100 bp
  *  118004  125508: contig of 7505 bp in length
  *  125509  125608: gap of 100 bp
  *  125609  137315: contig of 11707 bp in length
  *  137316  137415: gap of 100 bp
  *  137416  148418: contig of 11003 bp in length
  *  148419  148518: gap of 100 bp
  *  148519  161620: contig of 13102 bp in length
  *  161621  161720: gap of 100 bp
  *  161721  177868: contig of 16148 bp in length
```

```
RESULT 29
LOCUS       AP001899    183412 bp    DNA             HTG       26-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-70IC7 map 18q12, WORKING
            DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION   AP001899
VERSION     AP001899.3  GI:9501833
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-701C7.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183412)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 183,412 genomic DNA of 18q12
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 183412)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
```

```
Qy      593  ggtggttgccccaaccagc  612
             |||||||||||||||||||
Db    57383  GGTGGTTGCCCCAACCAGC  57402
```

```
Query Match             2.0%;  Score 20;  DB 69;  Length 177868;
Best Local Similarity   100.0%;  Pred. No. 7.9;
Matches     20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
AP001899/c
AP001899
```

JOURNAL   Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
            On Jul 26, 2000 this sequence version replaced gi:8117550.
COMMENT   ---------------------------------------
            Genome Center

              Center: RIKEN Genomic Sciences Center(GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@gsc.riken.go.jp

              Project Information

              Center project name: HumDraft18
              Center clone name: RP11-701C7

              Summary Statistics

              Sequencing vector: PCR products; 100% of reads
              Chemistry: Dye-terminator ET-amersham; 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 180400 bases at least Q40
              Consensus quality: 181718 bases at least Q30
              Consensus quality: 182304 bases at least Q20
              Insert size: 182712; sum-of-contigs
              Quality coverage: 10.87x in Q20 bases; sum-of-contigs

            NOTE: This is a 'working draft' sequence. It currently consists of
            8 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated and the accession number will be
            as soon as it is available and the accession number will be
            preserved.
                1   42740 contig of 42740 bp in length
             42841   79938 contig of 37098 bp in length
             80039 109211 contig of 29173 bp in length
            109312 135418 contig of 26107 bp in length
            135519 161816 contig of 26298 bp in length
            161917 174638 contig of 12722 bp in length
            174739 180277 contig of   5539 bp in length
            180378 183412 contig of   3035 bp in length.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 8 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
                1   42740: contig of 42740 bp in length
             42741   42840: gap of   100 bp
             42841   79938: contig of 37098 bp in length
             79939   80038: gap of   100 bp
             80039 109211: contig of 29173 bp in length
            109212 109311: gap of   100 bp
            109312 135418: contig of 26107 bp in length
            135419 135518: gap of   100 bp
            135519 161816: contig of 26298 bp in length
            161817 161916: gap of   100 bp
            161917 174638: contig of 12722 bp in length
            174639 174738: gap of   100 bp
            174739 180277: contig of   5539 bp in length
            180278 180377: gap of   100 bp
            180378 183412: contig of   3035 bp in length.
FEATURES   Location/Qualifiers
     source   1..183412
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q12"
                     /clone="RP11-701C7"
     misc_feature   1..42740
                     /note="assembly_fragment"
     misc_feature   42841..79938
                     /note="assembly_fragment clone_end:SP6 vector_side:right"

     misc_feature   80039..109211
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature   109312..135418
                     /note="assembly_fragment"
     misc_feature   135519..161816
                     /note="assembly_fragment"
     misc_feature   161917..174638
                     /note="assembly_fragment"
     misc_feature   174739..180277
                     /note="assembly_fragment"
     misc_feature   180378..183412
                     /note="assembly_fragment"
BASE COUNT   49688 a 39575 c 41016 g 52433 t   700 others
ORIGIN

Query Match   2.0%;   Score 20; DB 83; Length 183412;
Best Local Similarity   100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   211 gctgcagcttcccaacttcc 230
            ||||||||||||||||||||
Db   121048 GCTGCAGCTTCCCAACTTCC 121029

RESULT 30
ACC009591
LOCUS   AC009591   183855 bp   DNA   HTG   26-MAY-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-335K21 map 4, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC009591
VERSION   AC009591.3 GI:7107764
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 183855)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE   1 (bases 1 to 183855)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 4, clone RP11-335K21
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 183855)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,M.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 28, 2000 this sequence version replaced gi:6094611.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----
            Genome Center

              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ----
              Project Information

              Center project name: L1744
              Center clone name: 335_K_21

              Summary Statistics

RESULT 31

LOCUS   AC022446   184855 bp   DNA   HTG   06-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone RP11-5H13, WORKING DRAFT SEQUENCE,
            10 unordered pieces.
ACCESSION   AC022446

First record

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-primer-amersham; 2% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158308 bases at least Q40
Consensus quality: 170972 bases at least Q30
Consensus quality: 177434 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 183355; sum-of-contigs
Quality coverage: 5.2 in Q20 bases; agarose-fp
Quality coverage: 5.4 in Q20 bas.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*   1   1031: contig of 1031 bp in length
* 1032 1131: gap of 100 bp
* 1132 2356: contig of 1225 bp in length
* 2357 2456: gap of 100 bp
* 2457 3536: contig of 1080 bp in length
* 3537 3636: gap of 100 bp
* 3637 8344: contig of 4708 bp in length
* 8345 8444: gap of 100 bp
* 8445 45297: contig of 36853 bp in length
* 45298 45397: gap of 100 bp
* 45398 183855: contig of 138458 bp in length.

FEATURES   Location/Qualifiers
     source   1..183855
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-335K21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..1031
                     /note="assembly_fragment"
     misc_feature   1132..2356
                     /note="assembly_fragment"
     misc_feature   2457..3536
                     /note="assembly_fragment"
     misc_feature   3637..8344
                     /note="assembly_fragment"
     misc_feature   8445..45297
                     /note="assembly_fragment"
     misc_feature   45398..183855
                     /note="assembly_fragment"
                     /clone_end:T7
                     vector_side:right"
                     /clone_end:SP6
                     vector_side:left"

BASE COUNT 62884 a 32653 c 31457 g 56345 t 516 others
ORIGIN

Qy   593 ggtggttgccccaaccagc 612
           |||||||||||||||||||
Db 61148 GGTGGTTGCCCCAACCAGC 61167

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Best Local Similarity 100.0%; Pred. No. 7.9;
Query Match 593 2.0%; Score 20; DB 61; Length 183855;

Second record

VERSION   AC022446.3 GI:7711748
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 184855)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 184855)
  AUTHORS   DOE Joint Genome Institute.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On May 6, 2000 this sequence version replaced gi:7341552.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 408466
            Center clone name: RPCI-11_5H13

            Summary Statistics
            Consensus quality: 173751 bases at least Q40
            Consensus quality: 180112 bases at least Q30
            Consensus quality: 181210 bases at least Q20
            Estimated insert size: 194880; agarose-fp estimation
            Estimated insert size: 183955; sum-of-contigs estimation
            Quality coverage: 5.29 in Q20 bases; agarose-fp estimation
            Quality coverage: 5.6 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   4597: contig of 4597 bp in length
            * 4598   4697: gap of unknown length
            * 4698   11489: contig of 6792 bp in length
            * 11490   11589: gap of unknown length
            * 11590   16331: contig of 4742 bp in length
            * 16332   16431: gap of unknown length
            * 16432   24922: contig of 8491 bp in length
            * 24923   25022: gap of unknown length
            * 25023   38500: contig of 13478 bp in length
            * 38501   38600: gap of unknown length
            * 38601   57517: contig of 18917 bp in length
            * 57518   57617: gap of unknown length
            * 57618   78268: contig of 20651 bp in length
            * 78269   78368: gap of unknown length
            * 78369 103425: contig of 25057 bp in length
            * 103426 103525: gap of unknown length
            * 103526 135753: contig of 32228 bp in length
            * 135754 135853: gap of unknown length
            * 135854 184855: contig of 49002 bp in length.
FEATURES   Location/Qualifiers
     source   1..184855
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-5H13"
                     /clone_lib="RPCI human BAC library 11"

BASE COUNT   55073 a 36889 c 37015 g 54952 t 926 others
ORIGIN

Query Match 2.0%; Score 20; DB 67; Length 184855;
Best Local Similarity 100.0%; Pred. No. 7.9;

Qy   271 cccagcgcccagcagagaaaa 290
            |||||||||||||||||||||
Db   182417 CCCAGCCCCAGCAGAGAAAA 182436

Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 32
AC090095
AC090095   192059 bp   DNA   HTG   16-MAR-2001
LOCUS   Homo sapiens chromosome 8 clone RP11-296C13 map 8, WORKING DRAFT
DEFINITION SEQUENCE, 23 unordered pieces.
ACCESSION   AC090095
VERSION   AC090095.2 GI:13357359
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 192059)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens chromosome 8, clone RP11-296C13
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 192059)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Cah111,P., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,I., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (14-FEB-2001) Whitehead Institute/ MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 16, 2001 this sequence version replaced gi:12830252.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------------- Project Information
            Center project name: L12760
            Center clone name: 296_C_13
            -------------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 183542 bases at least Q40
            Consensus quality: 187455 bases at least Q30
            Consensus quality: 188948 bases at least Q20
            Insert size: 189859; sum-of-contigs
            Quality coverage: 4.5 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1   957: contig of 957 bp in length
            * 958   1057: gap of 100 bp
            * 1058   2239: contig of 1182 bp in length
            * 2240   2339: gap of 100 bp
            * 2340   3382: contig of 1043 bp in length
            * 3383   3482: gap of 100 bp
            * 3483   4503: contig of 1021 bp in length
            * 4504   4603: gap of 100 bp
            * 4604   7524: contig of 2921 bp in length
            * 7525   7624: gap of 100 bp
            * 7625   9358: contig of 1734 bp in length
            * 9359   9458: gap of 100 bp
            * 9459   12163: contig of 2705 bp in length
            * 12164   12263: gap of 100 bp
            * 12264   16386: contig of 4123 bp in length
            * 16387   16486: gap of 100 bp
            * 16487   19796: contig of 3310 bp in length
            * 19797   19896: gap of 100 bp
            * 19897   22758: contig of 2862 bp in length
            * 22759   22858: gap of 100 bp
            * 22859   26552: contig of 3694 bp in length
            * 26553   26652: gap of 100 bp
            * 26653   28501: contig of 1849 bp in length
            * 28502   28601: gap of 100 bp
            * 28602   32930: contig of 4329 bp in length
            * 32931   33030: gap of 100 bp
            * 33031   39585: contig of 6555 bp in length
            * 39586   39685: gap of 100 bp
            * 39686   46281: contig of 6596 bp in length
            * 46282   46381: gap of 100 bp
            * 46382   52654: contig of 6273 bp in length
            * 52655   52754: gap of 100 bp
            * 52755   57436: contig of 4682 bp in length
            * 57437   57536: gap of 100 bp
            * 57537   64278: contig of 6742 bp in length
            * 64279   64378: gap of 100 bp
            * 64379   73637: contig of 9259 bp in length
            * 73638   73737: gap of 100 bp
            * 73738   87928: contig of 14191 bp in length
            * 87929   88028: gap of 100 bp
            * 88029   106571: contig of 18543 bp in length
            * 106572 106671: gap of 100 bp
            * 106672 158839: contig of 52168 bp in length
            * 158840 158939: gap of 100 bp
            * 158940 192059: contig of 33120 bp in length

FEATURES   Location/Qualifiers
     source   1..192059
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-296C13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..957
                     /note="assembly_fragment
                     vector_side:left"
                     clone_end:SP6
     misc_feature   1058..2239
                     /note="assembly_fragment"
     misc_feature   2340..3382
                     /note="assembly_fragment"
     misc_feature   3483..4503
                     /note="assembly_fragment"
     misc_feature   4604..7524
                     /note="assembly_fragment"
     misc_feature   7625..9358
                     /note="assembly_fragment"
     misc_feature   9459..12163

RESULT 33

```
Qy        888 atgggaaagatgccacct 907
              ||||||||||||||||||
Db     118005 ATGGGGAAAGATGGCCACCT 118024
```

Query Match   2.08;   Score 20; DB 77;   Length 192059;
Best Local Similarity   100.0%;   Pred. No. 7.9;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

BASE COUNT   62894 a   33656 c   33229 g   60074 t   2206 others
ORIGIN

misc_feature   12264..16386
                /note="assembly_fragment"
misc_feature   16487..19796
                /note="assembly_fragment"
misc_feature   19897..22258
                /note="assembly_fragment"
misc_feature   22859..26552
                /note="assembly_fragment"
misc_feature   26653..28501
                /note="assembly_fragment"
misc_feature   28602..32930
                /note="assembly_fragment"
misc_feature   33031..35585
                /note="assembly_fragment"
misc_feature   39686..46281
                /note="assembly_fragment"
misc_feature   46382..52654
                /note="assembly_fragment"
misc_feature   52755..57436
                /note="assembly_fragment"
misc_feature   57537..64278
                /note="assembly_fragment"
misc_feature   64379..73637
                /note="assembly_fragment"
misc_feature   73738..87928
                /note="assembly_fragment"
misc_feature   88029..106571
                /note="assembly_fragment"
misc_feature   106672..158839
                /note="assembly_fragment"
misc_feature   158940..192059
                /note="assembly_fragment"
                /note="assembly_fragment"
                /note="clone_end:T7
                vector_side:right"

LOCUS   AC011804   195383 bp   DNA   HTG   03-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-71F23 map 18, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC011804
VERSION   AC011804.3 GI:8901273
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 195383)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 18, clone RP11-71F23
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 195383)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavky,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,J.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jul 3, 2000 this sequence version replaced gi:7329491.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L963
            Center clone name: 71_F23
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 95% of reads
            Sequencing vector: Plasmid; n/a; 8-0.f%% of reads
            4.842342342334234Chemistry: Dye-primer-amersham; 5% of reads
            Chemistry: Dye-terminator Big Dye; 95% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 174484 bases at least Q40
            Consensus quality: 183711 bases at least Q30
            Consensus quality: 187497 bases at least Q20
            Insert size: 200000; agarose-fp
            Insert size: 192383; sum-of-contigs
            Quality coverage:
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1875: contig of 1875 bp in length
            *   1876   1975: gap   of 100 bp
            *   1976   3161: contig of 1186 bp in length
            *   3162   3261: gap   of 100 bp
            *   3262   4663: contig of 1402 bp in length
            *   4664   4763: gap   of 100 bp
            *   4764   6577: contig of 1814 bp in length
            *   6578   6677: gap   of 100 bp
            *   6678   8647: contig of 1970 bp in length
            *   8648   8747: gap   of 100 bp
            *   8748   10746: contig of 1999 bp in length
            * 10747   10846: gap   of 100 bp
            * 10847   12332: contig of 1486 bp in length
            * 12333   12432: gap   of 100 bp
            * 12433   15508: contig of 3076 bp in length
            * 15509   15608: gap   of 100 bp
            * 15609   18286: contig of 2678 bp in length
            * 18287   18386: gap   of 100 bp
            * 18387   20319: contig of 1933 bp in length
            * 20320   20419: gap   of 100 bp
            * 20420   23111: contig of 2692 bp in length
            * 23112   23211: gap   of 100 bp
            * 23212   26610: contig of 3399 bp in length
            * 26611   26710: gap   of 100 bp
            * 26711   30644: contig of 3934 bp in length
            * 30645   30744: gap   of 100 bp
            * 30745   36084: contig of 5340 bp in length
            * 36085   36184: gap   of 100 bp
            * 36185   39659: contig of 3475 bp in length
            * 39660   39759: gap   of 100 bp
            * 39760   45091: contig of 5332 bp in length
            * 45092   45191: gap   of 100 bp in length
```

```
Qy    211 gtcgcagcttcccaacttcc 230
          ||||||||||||||||||||
Db  25633 GCTGCAGCTTCCCAACTTCC 25652
```

BASE COUNT   52065 a   43628 c   42262 g   54407 t   3021 others

```
Query Match         2.0%;    Score  20;  DB 62;  Length 195383;
Best Local Similarity  100.0%;  Pred. No. 7.9;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

RESULT 34
AL358492/c
```
LOCUS       AL358492   197788 bp    DNA             HTG       09-MAR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-340B24, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL358492
VERSION     AL358492.9  GI:13274738
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197788)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Mar 12, 2001 this sequence version replaced gi:13273744.
COMMENT     -------------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------------- Project Information
            Center project name: bA340B24
            -------------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 194817 bases at least Q40
            Consensus quality: 195716 bases at least Q30
```

```
FEATURES             Location/Qualifiers
     source          1..195383
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-71F23"
     misc_feature    1..1875
                     /note="assembly_fragment"
     misc_feature    1976..3161
                     /note="assembly_fragment"
     misc_feature    3262..4663
                     /note="assembly_fragment"
     misc_feature    4764..6577
                     /note="assembly_fragment"
     misc_feature    6678..8647
                     /note="assembly_fragment"
     misc_feature    8748..10746
                     /note="assembly_fragment"
     misc_feature    10847..12332
                     /note="assembly_fragment"
     misc_feature    12433..15508
                     /note="assembly_fragment"
     misc_feature    15609..18286
                     /note="assembly_fragment"
     misc_feature    18387..20319
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:right"
     misc_feature    20420..23111
                     /note="assembly_fragment"
     misc_feature    23212..26610
                     /note="assembly_fragment"
     misc_feature    26711..30644
                     /note="assembly_fragment"
     misc_feature    30745..36084
                     /note="assembly_fragment"
     misc_feature    36185..39659
                     /note="assembly_fragment"
     misc_feature    39760..45091
                     /note="assembly_fragment"
     misc_feature    45192..50472
                     /note="assembly_fragment"
     misc_feature    50573..55996
                     /note="assembly_fragment"
     misc_feature    56097..62551
                     /note="assembly_fragment"
     misc_feature    62652..67720
                     /note="assembly_fragment"
     misc_feature    67821..74113
                     /note="assembly_fragment"
     misc_feature    74214..80371
                     /note="assembly_fragment"
     misc_feature    80472..85618
                     /note="assembly_fragment"
     misc_feature    85719..96060
                     /note="assembly_fragment"
     misc_feature    96161..103014
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:left"
     misc_feature    103115..114047
                     /note="assembly_fragment"
     misc_feature    114148..125353
                     /note="assembly_fragment"
     misc_feature    125454..137602
                     /note="assembly_fragment"
     misc_feature    137703..150975
                     /note="assembly_fragment"
     misc_feature    151076..173481
                     /note="assembly_fragment"
     misc_feature    173582..195383
                     /note="assembly_fragment"
```

```
*      45192   50472: gap of 100 bp
       50473   50572: gap of 100 bp
       50573   55996: contig of 5424 bp in length
       55997   56096: gap of 100 bp
       56097   62551: contig of 6455 bp in length
       62552   62651: gap of 100 bp
       62652   67720: contig of 5069 bp in length
       67721   67820: gap of 100 bp
       67821   74113: contig of 6293 bp in length
       74114   74213: gap of 100 bp
       74214   80371: contig of 6158 bp in length
       80372   80471: gap of 100 bp
       80472   85618: contig of 5147 bp in length
       85619   85718: gap of 100 bp
       85719   96060: contig of 10342 bp in length
       96061   96160: gap of 100 bp
       96161  103014: contig of 6854 bp in length
      103015  103114: gap of 100 bp
      103115  114047: contig of 10933 bp in length
      114048  114147: gap of 100 bp
      114148  125353: contig of 11206 bp in length
      125354  125453: gap of 100 bp
      125454  137602: contig of 12149 bp in length
      137603  137702: gap of 100 bp
      137703  150975: contig of 13273 bp in length
      150976  151075: gap of 100 bp
      151076  173481: contig of 22406 bp in length
      173482  173581: gap of 100 bp
      173582  195383: contig of 21802 bp in length.
```

Matches   20;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     10 caagattcaaacccagctca 29
          ||||||||||||||||||||
Db  15096 CAAGATTCAAACCCAGCTCA 15077
```

RESULT 35

```
LOCUS       AC018443   199812 bp   DNA            HTG        26-FEB-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-674A24 map 8, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC018443
VERSION     AC018443.6  GI:13123355
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199812)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-674A24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199812)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W.,
            Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,
            Heaford,A., Horton,L., Howland,J., Johnson,R., Jones,C., Kann,L.,
            Karatas,A., Klein,J., Landers,T., Lehoczky,J., Lieu,C., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., Peterson,K., Pierce,N., Pollara,V., Riley,R.,
            Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 25, 2001 this sequence version replaced gi:7249139.
            All repeats were identified using RepeatMasker;
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L5117
            Center clone name: 674_A_24
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 63% of reads
            Sequencing vector: Plasmid; n/a; 37% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 198051 bases at least Q40
            Consensus quality: 198846 bases at least Q30
            Consensus quality: 199187 bases at least Q20
            Insert size: 194000; agarose-fp
            Insert size: 199412; sum-of-contigs
            Quality coverage: 7.7 in Q20 bases; agarose-fp
            Quality coverage: 7.5 in Q20 b.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
Consensus quality: 196299 bases at least Q20
Insert size: 196988; sum-of-contigs
Insert size: 185634; 10.0% error; agarose-fp
Quality coverage: 7.00x in Q20 bases; sum-of-contigs Quality
coverage: 7.67x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

    1  29856: contig of 29856 bp in length
29857  29956: gap of 100 bp in length
29957  56368: contig of 26412 bp in length
56369  56468: gap of 100 bp in length
56469  66708: contig of 10240 bp in length
66709  66808: gap of 100 bp in length
66809  84676: contig of 17868 bp in length
84677  84776: gap of 100 bp in length
84777 121388: contig of 36612 bp in length
121389 121488: gap of 100 bp in length
121489 128913: contig of 7425 bp in length
128914 129013: gap of 100 bp in length
129014 162549: contig of 33536 bp in length
162550 162649: gap of 100 bp in length
162650 185166: contig of 22517 bp in length
185167 185266: gap of 100 bp in length
185267 197788: contig of 12522 bp in length.

FEATURES             Location/Qualifiers
     source          1..197788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-340B24"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..29856
                     /note="assembly_fragment:00892"
     misc_feature    29957..56368
                     /note="assembly_fragment:02882"
     misc_feature    56469..66708
                     /note="assembly_fragment:01981"
     misc_feature    66809..84676
                     /note="assembly_fragment:02337"
     misc_feature    84777..121388
                     /note="assembly_fragment:00717"
     misc_feature    121489..128913
                     /note="assembly_fragment:03116"
     misc_feature    129014..162549
                     /note="assembly_fragment:00061"
     misc_feature    162650..185166
                     /note="assembly_fragment:01279"
     misc_feature    185267..197788
                     /note="assembly_fragment:00238"

BASE COUNT     45671 a    51501 c    50781 g    49025 t    810 others
ORIGIN
```

Query Match   2.0%;   Score 20;   DB 80;   Length 197788;
Best Local Similarity   100.0%;   Pred. No. 7.9;

```
RESULT 36

Qy    888 atggggaaagatgccaccct 907
          ||||||||||||||||||||
Db  87641 ATGGGGGAAAGATGCCACCCT 87660

Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
Best Local Similarity 100.0%;  Pred. No. 7.9;
Query Match 2.0%;  Score 20;  DB 65;  Length 199812;
```

```
FEATURES             Location/Qualifiers
     source          1..199812
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-674A24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..15366
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15467..36444
                     /note="assembly_fragment"
     misc_feature    36545..107322
                     /note="assembly_fragment"
     misc_feature    107423..162742
                     /note="assembly_fragment"
     misc_feature    162843..199812
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:right"
BASE COUNT     66339 a   34885 c   34539 g   63649 t   400 others
ORIGIN
```

```
*                           be preserved.
*        1    15366: contig of 15366 bp in length
*    15367    15466: gap of 100 bp
*    15467    36444: contig of 20978 bp in length
*    36445    36544: gap of 100 bp
*    36545   107322: contig of 70778 bp in length
*   107323   107422: gap of 100 bp
*   107423   162742: contig of 55320 bp in length
*   162743   162842: gap of 100 bp
*   162843   199812: contig of 36970 bp in length.
```

```
LOCUS       AC068379    222876 bp    DNA    HTG    22-OCT-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-537I16 map 3, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC068379
VERSION     AC068379.4  GI:1094576
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222876)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-537I16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 222876)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Bouslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Cheepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Donino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 22, 2000 this sequence version replaced gi:10305231.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------  Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------------  Project Information
            Center project name: L10217
            Center clone name: 537_I16
            -------------------
            NOTE: This record contains 278 individual
            sequencing reads that have not been assembled into
            contigs. Runs of N are used to separate the reads
            and the order in which they appear is completely
            arbitrary. Low-pass sequence sampling is useful for
            identifying clones that may be gene-rich and allows
            overlap relationships among clones to be deduced.
            However, it should not be assumed that this clone
            will be sequenced to completion. In the event that
            the record is updated, the accession number will
            be preserved.
```

```
*        1      728: contig of 728 bp in length
*      729      828: gap of 100 bp
*      829     1559: contig of 731 bp in length
*     1560     1659: gap of 100 bp
*     1660     2382: contig of 723 bp in length
*     2383     2482: gap of 100 bp
*     2483     3231: contig of 749 bp in length
*     3232     3331: gap of 100 bp
*     3332     4056: contig of 725 bp in length
*     4057     4156: gap of 100 bp
*     4157     4881: contig of 725 bp in length
*     4882     4981: gap of 100 bp
*     4982     5714: contig of 733 bp in length
*     5715     5814: gap of 100 bp
*     5815     6526: contig of 712 bp in length
*     6527     6626: gap of 100 bp
*     6627     7343: contig of 717 bp in length
*     7344     7443: gap of 100 bp
*     7444     8173: contig of 730 bp in length
*     8174     8273: gap of 100 bp
*     8274     8980: contig of 707 bp in length
*     8981     9080: gap of 100 bp
*     9081     9788: contig of 708 bp in length
*     9789     9888: gap of 100 bp
*     9889    10620: contig of 732 bp in length
*    10621    10720: gap of 100 bp
*    10721    11471: contig of 751 bp in length
*    11472    11571: gap of 100 bp
*    11572    12317: contig of 746 bp in length
*    12318    12417: gap of 100 bp
*    12418    13148: contig of 731 bp in length
*    13149    13248: gap of 100 bp
*    13249    13973: contig of 725 bp in length
*    13974    14073: gap of 100 bp
*    14074    14812: contig of 739 bp in length
*    14813    14912: gap of 100 bp
```

```
*    14913   15633: contig of 721 bp in length
*    15634   15733: gap of 100 bp
*    15734   16475: contig of 742 bp in length
*    16476   16575: gap of 100 bp
*    16576   17284: contig of 709 bp in length
*    17285   17384: gap of 100 bp
*    17385   18078: contig of 694 bp in length
*    18079   18178: gap of 100 bp
*    18179   18888: contig of 710 bp in length
*    18889   18988: gap of 100 bp
*    18989   19732: contig of 744 bp in length
*    19733   19832: gap of 100 bp
*    19833   20559: contig of 727 bp in length
*    20560   20659: gap of 100 bp
*    20660   21391: contig of 732 bp in length
*    21392   21491: gap of 100 bp
*    21492   22218: contig of 727 bp in length
*    22219   22318: gap of 100 bp
*    22319   23037: contig of 719 bp in length
*    23038   23137: gap of 100 bp
*    23138   23846: contig of 709 bp in length
*    23847   23946: gap of 100 bp
*    23947   24668: contig of 722 bp in length
*    24669   24768: gap of 100 bp
*    24769   25500: contig of 732 bp in length
*    25501   25600: gap of 100 bp
*    25601   26321: contig of 721 bp in length
*    26322   26421: gap of 100 bp
*    26422   27131: contig of 710 bp in length
*    27132   27231: gap of 100 bp
*    27232   27952: contig of 721 bp in length
*    27953   28052: gap of 100 bp
*    28053   28736: contig of 684 bp in length
*    28737   28836: gap of 100 bp
*    28837   29582: contig of 746 bp in length
*    29583   29682: gap of 100 bp
*    29683   30397: contig of 715 bp in length
*    30398   30497: gap of 100 bp
*    30498   31242: contig of 745 bp in length
*    31243   31342: gap of 100 bp
*    31343   32069: contig of 727 bp in length
*    32070   32169: gap of 100 bp
*    32170   32889: contig of 720 bp in length
*    32890   32989: gap of 100 bp
*    32990   33723: contig of 734 bp in length
*    33724   33823: gap of 100 bp
*    33824   34523: contig of 700 bp in length
*    34524   34623: gap of 100 bp
*    34624   35343: contig of 719 bp in length
*    35343   35442: gap of 100 bp
*    35443   36172: contig of 730 bp in length
*    36173   36272: gap of 100 bp
*    36273   36988: contig of 716 bp in length
*    36989   37088: gap of 100 bp
*    37089   37820: contig of 732 bp in length
*    37821   37920: gap of 100 bp
*    37920   38636: contig of 716 bp in length
*    38637   38736: gap of 100 bp
*    38737   39455: contig of 719 bp in length
*    39456   39555: gap of 100 bp
*    39555   40248: contig of 694 bp in length
*    40249   40349: gap of 100 bp
*    40350   41074: contig of 725 bp in length
*    41075   41174: gap of 100 bp
*    41175   41907: contig of 733 bp in length
*    41908   42007: gap of 100 bp
*    42008   42726: contig of 719 bp in length
*    42727   42826: gap of 100 bp
*    42827   43579: contig of 753 bp in length
*    43580   43679: gap of 100 bp
*    43680   44374: contig of 695 bp in length
*    44375   44474: gap of 100 bp
*    44475   45180: contig of 706 bp in length
*    45181   45280: gap of 100 bp
*    45281   45979: contig of 699 bp in length
*    45980   46079: gap of 100 bp
*    46080   46821: contig of 742 bp in length
*    46822   46921: gap of 100 bp
*    46922   47645: contig of 724 bp in length
*    47646   47745: gap of 100 bp
*    47746   48465: contig of 720 bp in length
*    48466   48565: gap of 100 bp
*    48566   49289: contig of 724 bp in length
*    49290   49389: gap of 100 bp
*    49390   50107: contig of 718 bp in length
*    50108   50207: gap of 100 bp
*    50208   50925: contig of 718 bp in length
*    50926   51025: gap of 100 bp
*    51026   51736: contig of 711 bp in length
*    51737   51836: gap of 100 bp
*    51837   52543: contig of 707 bp in length
*    52544   52643: gap of 100 bp
*    52644   53346: contig of 703 bp in length
*    53347   53446: gap of 100 bp
*    53447   54192: contig of 746 bp in length
*    54193   54292: gap of 100 bp
*    54293   55014: contig of 722 bp in length
*    55015   55114: gap of 100 bp
*    55115   55824: contig of 710 bp in length
*    55825   55924: gap of 100 bp
*    55925   56651: contig of 727 bp in length
```

```
Query Match     2.0%;  Score 20;  DB 73;  Length 222876;
Best Local Similarity 100.0%;  Pred. No. 7.9;
Matches  20;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      449 gaccaggtggacagggtcct 468
            ||||||||||||||||||||
Db   106725 GACCAGGTGGACAGGGTCCT 106744
```

```
RESULT 37
AC068379/c
LOCUS       AC068379     222876 bp    DNA    HTG    22-OCT-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-537I16 map 3, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC068379
VERSION     AC068379.4  GI:10945763
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222876)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 3, clone RP11-537I16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 222876)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McKwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Nguyen,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
```

Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE   Direct Submission

JOURNAL   Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
          On Oct 22, 2000 this sequence version replaced gi:10305231.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
          --------- Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
          --------- Project Information
          Center project name: L10217
          Center clone name: 537_1_16

          * NOTE: This record contains 278 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.

          *   1   728: contig of 728 bp in length
          *   729   828: gap of 100 bp
          *   829   1559: contig of 731 bp in length
          *   1560   1659: gap of 100 bp
          *   1660   2382: contig of 723 bp in length
          *   2383   2482: gap of 100 bp
          *   2483   3231: contig of 749 bp in length
          *   3232   3331: gap of 100 bp
          *   3332   4056: contig of 725 bp in length
          *   4057   4156: gap of 100 bp
          *   4157   4881: contig of 725 bp in length
          *   4882   4981: gap of 100 bp
          *   4982   5714: contig of 733 bp in length
          *   5715   5814: gap of 100 bp
          *   5815   6526: contig of 712 bp in length
          *   6527   6626: gap of 100 bp
          *   6627   7343: contig of 717 bp in length
          *   7344   7443: gap of 100 bp
          *   7444   8173: contig of 730 bp in length
          *   8174   8273: gap of 100 bp
          *   8274   8980: contig of 707 bp in length
          *   8981   9080: gap of 100 bp
          *   9081   9788: contig of 708 bp in length
          *   9789   9888: gap of 100 bp
          *   9889 10620: contig of 732 bp in length
          *   10621 10720: gap of 100 bp
          *   10721 11471: contig of 751 bp in length
          *   11472 11571: gap of 100 bp
          *   11572 12317: contig of 746 bp in length
          *   12318 12417: gap of 100 bp
          *   12418 13148: contig of 731 bp in length
          *   13149 13248: gap of 100 bp
          *   13249 13973: contig of 725 bp in length
          *   13974 14073: gap of 100 bp
          *   14074 14812: contig of 739 bp in length
          *   14813 14912: gap of 100 bp
          *   14913 15633: contig of 721 bp in length
          *   15634 15733: gap of 100 bp
          *   15734 16475: contig of 742 bp in length
          *   16476 16575: gap of 100 bp
          *   16576 17284: contig of 709 bp in length
          *   17285 17384: gap of 100 bp
          *   17385 18078: contig of 694 bp in length

          *   18079 18178: gap of 100 bp in length
          *   18179 18888: contig of 710 bp in length
          *   18889 18988: gap of 100 bp
          *   18989 19732: contig of 744 bp in length
          *   19733 19832: gap of 100 bp
          *   19833 20559: contig of 727 bp in length
          *   20560 20659: gap of 100 bp
          *   20660 21391: contig of 732 bp in length
          *   21392 21491: gap of 100 bp
          *   21492 22218: contig of 727 bp in length
          *   22219 22318: gap of 100 bp
          *   22319 23037: contig of 719 bp in length
          *   23038 23137: gap of 100 bp
          *   23138 23846: contig of 709 bp in length
          *   23847 23946: gap of 100 bp
          *   23947 24668: contig of 722 bp in length
          *   24669 24768: gap of 100 bp
          *   24769 25500: contig of 732 bp in length
          *   25501 25600: gap of 100 bp
          *   25601 26321: contig of 721 bp in length
          *   26322 26421: gap of 100 bp
          *   26422 27131: contig of 710 bp in length
          *   27132 27231: gap of 100 bp
          *   27232 27952: contig of 721 bp in length
          *   27953 28052: gap of 100 bp
          *   28053 28736: contig of 684 bp in length
          *   28737 28836: gap of 100 bp
          *   28837 29582: contig of 746 bp in length
          *   29583 29682: gap of 100 bp
          *   29683 30397: contig of 715 bp in length
          *   30398 30497: gap of 100 bp
          *   30498 31242: contig of 745 bp in length
          *   31243 31342: gap of 100 bp
          *   31343 32069: contig of 727 bp in length
          *   32070 32169: gap of 100 bp
          *   32170 32889: contig of 720 bp in length
          *   32890 32989: gap of 100 bp
          *   32990 33723: contig of 734 bp in length
          *   33724 33823: gap of 100 bp
          *   33824 34523: contig of 700 bp in length
          *   34524 34623: gap of 100 bp
          *   34624 35342: contig of 719 bp in length
          *   35343 35442: gap of 100 bp
          *   35443 36172: contig of 730 bp in length
          *   36173 36272: gap of 100 bp
          *   36273 36988: contig of 716 bp in length
          *   36989 37088: gap of 100 bp
          *   37089 37820: contig of 732 bp in length
          *   37821 37920: gap of 100 bp
          *   37921 38636: contig of 716 bp in length
          *   38637 38736: gap of 100 bp
          *   38737 39455: contig of 719 bp in length
          *   39456 39555: gap of 100 bp
          *   39556 40249: contig of 694 bp in length
          *   40250 40349: gap of 100 bp
          *   40350 41074: contig of 725 bp in length
          *   41075 41174: gap of 100 bp
          *   41175 41907: contig of 733 bp in length
          *   41908 42007: gap of 100 bp
          *   42008 42726: contig of 719 bp in length
          *   42727 42826: gap of 100 bp
          *   42827 43579: contig of 753 bp in length
          *   43580 43679: gap of 100 bp
          *   43680 44374: contig of 695 bp in length
          *   44375 44474: gap of 100 bp
          *   44475 45180: contig of 706 bp in length
          *   45181 45280: gap of 100 bp
          *   45281 45979: contig of 699 bp in length
          *   45980 46079: gap of 100 bp
          *   46080 46821: contig of 742 bp in length
          *   46822 46921: gap of 100 bp
          *   46922 47645: contig of 724 bp in length
          *   47646 47745: gap of 100 bp

```
Qy   449 gaccaggtggacaggtcct 468
         ||||||||||||||||||
Db 222603 GACCAGGTGGACAGGTCCT 222584
```

```
RESULT 38
AC004615/c
LOCUS       AC004615   235141 bp    DNA              PRI        03-FEB-2000
DEFINITION  Homo sapiens BAC clone GS1-405L21 from 5p15.2, complete sequence.
ACCESSION   AC004615
VERSION     AC004615.1   GI:3080660
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 235141)
  AUTHORS   Kalicki,J. and Harmon,G.
  TITLE     The sequence of Homo sapiens BAC clone GS1-405L21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235141)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   3  (bases 1 to 235141)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            --------------  Summary Statistics
            Center project name: H_GS405L21
            --------------
            --------------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
```

```
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Michael
Lovett, Departments of Otorhinolaryngology, Molecular Biology and
Oncology, University of Texas Southwestern Medical Center, Dallas
TX. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc.
(http://www.genomesystems.com).
Cell line:   lymphoblastoid
Haplotypes:  two
VECTOR: pBeloBAC
Selection:   chloramphenicol
NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of GS1-405L21;
actual end is at 235141 of GS1-405L21.

This clone contains STS HSA0552D9 (NID:g1233025).
```

```
FEATURES             Location/Qualifiers
     source          1..235141
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p15.2"
                     /clone="GS1-405L21"
                     /clone_lib="GSBAC1"
     repeat_region   181..480
                     /rpt_family="Alu"
     repeat_region   3000..3303
                     /rpt_family="Alu"
     repeat_region   4356..4559
                     /rpt_family="Alu"
     repeat_region   4576..4744
                     /rpt_family="(TAGA)n"
     repeat_region   6769..7074
                     /rpt_family="MER1_type"
     repeat_region   7096..7157
                     /rpt_family="Alu"
     repeat_region   7286..7384
                     /rpt_family="(CA)n"
     repeat_region   7491..7742
                     /rpt_family="MaLR"
     repeat_region   9154..>234415
                     /rpt_family="MaLR"
     gene            9154..>234415
                     /gene="SEMAF"
     CDS             join(9154..9277,51204..51303,70597..70642,151152..151214,
                     162076..162174,164155..164368,186819..187104,
                     191786..191921,198574..198778,234207..>234415)
                     /gene="SEMAF"
                     /note="match to U52840 (PID:g2772584);but note
                     polymorphism at position 56; H_GS405L21.1"
                     /codon_start=1
                     /product="semaphorin F"
                     /protein_id="AAC14668.1"
                     /db_xref="GI:3080661"
                     /translation="MKGTCVIAWLFSSLGLWRLAHPEAQGTTQCQRTEHPVISYKEIG
                     PSFMEDNSRFSHVAVDVVGGREALVHIIYLATDYGTIKKVRPLNQTSSSCLLEEIEL
                     FPERRREPIRSLGILHSQSVLFVGLREHVKIPLKRCQFYRTRR"
                     PWLREFRRAKNAVDFSQLFDFQGKELVVGARNYLFRLOLEDLSLIQAVEWEDEATKK
                     ACTSKGSKKEECCNYIRVLLVGGDRLFTCGTNAFTPVCTNRSLSNLTEIHDQISGMAR
                     CPYSPQHNSTALLTAGGELYAAVTAMDFPGRDPAIYRSLGILPLRTAQYNSKWLNEPN
                     FVSSYDIGNFTYFFRENAVEHDCGKTVFSRAARVCKNDIGGRFLLEDTFMKARL
                     NCSRPGEVPFYYNELQSTFPELEDLIGIFTTNVVNSIAASAVCVFNLSAIAQAFSGP
                     FKYQENSRSAWLPYPNPNHFQCGTVDGLYVNLTERNLQDAQKFILMHEVVQPVTTV
```

```
repeat_region   10974..11242
                /rpt_family="L1"
repeat_region   11271..11381
                /rpt_family="L1"
repeat_region   12176..12392
```

repeat_region   13349..13391
                 /rpt_family="L1"
repeat_region   13453..13898
                 /rpt_family="MER81"
repeat_region   19001..19202
                 /rpt_family="Alu"
repeat_region   19350..19841
                 /rpt_family="MaLR"
repeat_region   19893..20700
                 /rpt_family="L2"
repeat_region   21002..21401
                 /rpt_family="L1"
repeat_region   21797..21994
                 /rpt_family="L1"
repeat_region   22413..22710
                 /rpt_family="MaLR"
repeat_region   27381..27411
                 /rpt_family="MER1_type"
repeat_region   27785..28082
                 /rpt_family="Alu"
repeat_region   29733..29893
                 /rpt_family="Mariner"
repeat_region   30444..30532
                 /rpt_family="MIR"
repeat_region   33326..33653
                 /rpt_family="L2"
repeat_region   33961..34048
                 /rpt_family="L2"
repeat_region   34057..34316
                 /rpt_family="Alu"
repeat_region   34390..34520
                 /rpt_family="MER1_type"
repeat_region   36072..36286
                 /rpt_family="L2"
repeat_region   36464..36604
                 /rpt_family="MER1_type"
repeat_region   36622..36732
                 /rpt_family="Alu"
repeat_region   38061..38641
                 /rpt_family="(TAGA)n"
repeat_region   38939..39082
                 /rpt_family="MER21_g"
repeat_region   39096..39399
                 /rpt_family="MER1_type"
repeat_region   39415..39483
                 /rpt_family="Alu"
repeat_region   39557..39683
                 /rpt_family="MER1_type"
repeat_region   42462..42767
                 /rpt_family="L2"
repeat_region   42862..43091
                 /rpt_family="Retroviral"
repeat_region   46106..46225
                 /rpt_family="Retroviral"
repeat_region   47785..47839
                 /rpt_family="MIR"
repeat_region   51495..51731
                 /rpt_family="(CA)n"
repeat_region   53061..53183
                 /rpt_family="MIR"
repeat_region   54138..54238
                 /rpt_family="L2"
repeat_region   54275..54314
                 /rpt_family="MER53"
repeat_region   54348..54495
                 /rpt_family="L2"
repeat_region   58611..58868
                 /rpt_family="MIR"
repeat_region   59061..59261
                 /rpt_family="Alu"
repeat_region   59469..59532
                 /rpt_family="MER2_type"
repeat_region   59571..59768
                 /rpt_family="L2"

repeat_region   60224..60521
                 /rpt_family="MER1_type"
repeat_region   62226..62479
                 /rpt_family="Alu"
repeat_region   62480..62757
                 /rpt_family="Alu"
repeat_region   63954..64048
                 /rpt_family="Alu"
repeat_region   64650..64950
                 /rpt_family="MIR"
repeat_region   65069..65357
                 /rpt_family="Alu"
                 /rpt_family="Alu"

Query Match   2.0%;   Score 20; DB 85; Length 235141;
Best Local Similarity   100.0%; Pred. No. 7.9;
Matches   20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   271 cccagccccagcagagaaa   290
              ||||||||||||||||||||
Db   156934 CCCAGCCCCAGCAGAGAAAA   156915

RESULT 39
AF178954/c
LOCUS   AF178954   2051 bp   mRNA   ROD   19-OCT-2000
DEFINITION   Mus musculus monocarboxylate transporter 4 mRNA, complete cds.
ACCESSION   AF178954
VERSION   AF178954.1   GI:10880481
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2051)
  AUTHORS   Yoon,H. and Philp,N.J.
  TITLE   Cloning and expression of mouse MCT3 and MCT4
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2051)
  AUTHORS   Yoon,H. and Philp,N.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-AUG-1999) Cell and Molecular Biology, Pennsylvania
              College of Optometry, 8360 Old York Road, Philadelphia, PA 19027,
              USA
FEATURES   Location/Qualifiers
     source   1..2051
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /tissue_type="skeletal muscle"
     CDS   1..1413
                     /note="MCT4"
                     /codon_start=1
                     /product="monocarboxylate transporter 4"
                     /protein_id="AAC24271.1"
                     /db_xref="GI:10880482"
                     /translation="MGGAVVDEGPTGIKAPDGGWGWAVLFGCFIITGFSVAFPKAVSV
                     FFKELMHEFGTGYSDTAWISSTLLAMLYGTGPLCSVCVNRFGCRPVMLVGGLFASLGM
                     VAASFCRSIIQIYLTTGVITGLGLALNFQPSLIMLNRYFNKRRPIANGLAAAGSPVFL
                     CALSPLGQLLQDHYGWRGGFFILGGLLLNCCVCAALMRPLVAPQVGGGTEPRGPQRPP
                     QRLLDLSVFRDRGFLIYAVAASIMVLGLFVPPVFVVSYAKDMGVPDTKAAFLLTIIGF
                     IDIFARPTAGFITGLKKVRPYKSVYLFSFAMFENGFTDLGTGSTATDYGGLIVFCIFGI
                     SYGMVGALQFEVLMAIVGTOKFSSAIGLVLLLEAVAVLIGPPSGGKLLLDATKVYKYVF
                     ILAGAEVLTSSLVLLLGNFFCIGKRRPEVTEPEEVASEEKLHKPPVDVGVDSREVEH
                     FLKAEPEKNGEVVHTPETSV"
BASE COUNT   602 a   608 c   469 g   372 t
ORIGIN

Query Match   1.9%;   Score 19; DB 94; Length 2051;
Best Local Similarity   100.0%; Pred. No. 30;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 40

```
Qy          264 aggacggcccagcccccagc 282
                ||||||||||||||||||||
Db           74 AGGACGGCCCAGCCCCCAGC 56
```

```
LOCUS       RNU87627                2118 bp    mRNA                    ROD       24-JUN-1998
DEFINITION  Rattus norvegicus putative monocarboxylate transporter (MCT3) mRNA,
            complete cds.
ACCESSION   U87627
VERSION     U87627.1  GI:2463650
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2118)
  AUTHORS   Wilson,M.C., Jackson,V.N., Heddle,C., Price,N.T., Pilegaard,H.,
            Juel,C., Bonen,A., Montgomery,I., Hutter,O.F. and Halestrap,A.P.
  TITLE     Lactic acid efflux from white skeletal muscle is catalyzed by the
            monocarboxylate transporter isoform MCT3
  JOURNAL   J. Biol. Chem. 273 (26), 15920-15926 (1998)
  MEDLINE   98298091
REFERENCE   2  (bases 1 to 2118)
  AUTHORS   Jackson,V.N., Price,N.T. and Halestrap,A.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1997) Cellular Biochemistry, Hannah Research
            Institute, Mauchlin Road, Ayr KA6 5HL, UK
FEATURES             Location/Qualifiers
     source          1..2118
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /dev_stage="12 weeks"
                     /sex="male"
                     /tissue_type="skeletal muscle"
     gene            1..2118
                     /gene="MCT3"
     CDS             90..1505
                     /gene="MCT3"
                     /note="similar to Gallus gallus retinal epithelial
                     membrane protein, encoded by GenBank Accession Number
                     U15685; similar to Human MCT3, encoded by Genbank
                     Accession Number U81800"
                     /codon_start=1
                     /product="putative monocarboxylate transporter"
                     /protein_id="AAC53591.1"
                     /db_xref="GI:2463651"
                     /translation="MGGAVVDEGPTGIKAPDGGWGWAVLFGCFITGFSYAFPKAVSV
                     FKELMHEFGTGYSDTAWISSILAMLYGTGPLCSMCVNRFGCRPVMLVGGLFASLGM
                     VAASFCRSIIQIYLTTGVITGLGLAINFQPSLIMLNRYFNRRPMANGLAAAGSPVPL
                     CALSPLGQLLQDHYGWRGGFLILGGLLLNCCVCAALMRPLVAPQASGGAEPHGPQRPS
                     PRLDLSVFRDRGFLIYAVAASIMVLGLFVPPVFVVSYAKDMGVPDTKAAFLLTLGF
                     IDIFARPTAGFITGLKKVRPYSVYLFSFAMFENGFTDLTGSTPASDYGGLVVPCIFFGI
                     SYGMVGALQEEVALMAIVGTQKFSSAIGLVLLLEAVAVLIGPPSGGKLLDATKVYKVF
                     ILAGAEVLTSSLVLLLGNFCIGKRRPEVTKPEEVASEEEKLHKPPVDVRDSREVE
                     HFLKAEPEKNGEVVHPETSV"
BASE COUNT      401 a    605 c    629 g    483 t
ORIGIN
```

```
Qy          264 aggacggcccagcccccagc 282
                ||||||||||||||||||||
Db          163 AGGACGGCCCAGCCCCCAGC 145
```

Query Match   1.98;   Score 19;   DB 95;   Length 2118;
Best Local Similarity   100.0%;   Pred. No. 30;
Matches   19;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 41

```
AF136600/c
LOCUS       AF136600                5841 bp    DNA                     PLN       01-JUN-1999
DEFINITION  Magnaporthe grisea protein kinase C (mpkc) gene, complete cds.
ACCESSION   AF136600
VERSION     AF136600.1  GI:4928704
KEYWORDS    .
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 5841)
  AUTHORS   Khang,C.H. and Lee,Y.H.
  TITLE     The identification and characterization of protein kinase C, mpkc,
            in Magnaporthe grisea
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5841)
  AUTHORS   Khang,C.H. and Lee,Y.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1999) Division of Applied Biology and Chemistry,
            Seoul National University, 103, Seodun-dong, Kwonsun-gu, Suwon 441
            - 744, South Korea
FEATURES             Location/Qualifiers
     source          1..5841
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /note="submitted through Biological Research Information
                     Center of Korea GeneNuri No. KS101751"
     gene            <883..>4777
                     /gene="mpkc"
     CDS             join(883..1342,1427..1693,1772..3923,3987..4188,
                     4242..4647,4716..4777)
                     /gene="mpkc"
                     /codon_start=1
                     /product="protein kinase C"
                     /protein_id="AAD33693.1"
                     /db_xref="GI:4928705"
                     /translation="MDDRIQDIYKIEEEKALLNAANLMRQQINNDAVRSKLDTQMRE
                     GRNLEFFEGTLREMQMRSMGQGMDNLSTGGSTLAASASSSSRPRSAVEEDGPMPPK
                     DGGYGAGGDGYGQTYQSQLGEHGDLMPPGRPFANQGPGCSHIPKTRPNFTKLDLIKYDT
                     PYLGPROIMLSQIOFKLNVEEQYLKGTEKMVOLYQMEGDKSKRADAAARVRESKQKI
                     VLIKQALRYEELHIDIDSGDSPDDIIIMPNLRKPLTGQLSIRVIAIKDVDHAPTGR
                     FARQPDTEVAIKVEDNIAARTRVSRTDRWENEYHNVEVDKANIELTVYDKPGEHSIP
                     IGMLVRISDIAEELRKRKIEAETNSLGWVSADRMGDVNPAVAVRPPQFPWNIENGP
                     GGPVSPGIQGOOGYGPCGPDGGCYGPGDPGVGEDPGVGEPTGWENELETIGSIQLSLNFFKO
                     SKDRRPVDLGLGRKGAIRQRKEEVHEMYCHKEVERQFYNIMRCALCGDFLKGGSGMQC
                     EDCKYTHCKCTSVTRCISKSNAEFDPDEEKINHRIPHRFVAFSNLTANWCCHGY
                     LPFGKKNCRKCSECOMAAHAGCVHLVPDFCGMSMAVANQILEGISROKQRONKOASSM
                     SDRFLRSGKMSPPGSGHASSASFSOGMGSSYGOASPEATEAAKMYSSQTSPORITSP
                     DRRSSSSOAAAAATAAMTGLPGAMSSOPGROOGOOPITDYGSASGGRVGSVGPHDDPYA
                     OPQSPPPOOOAAYGPQOORKYNPADYANISGCYGSOPQMAOOPOOARPOOOOOPLI
                     SPOOHASOAOOPLSPVKOOHOEOQIISPTAGTVIPTSARPLPSATDPGTGORIGLDH
                     ENFLAVLGGNGKVMLAESKKRTKRKLYAIKVLKKFEEIENDEVESIRSEKRVLLIANR
                     EHPFLTNLHACFQTETRVYFVEEYISGGDLMLHIQCFGTRAQFYAAEVCLALKY
                     FHENGVIYRDLKLDNILLTLDGHIKIADYGLCKEDMVGSTTSFCGTPEFMAPEILL
                     DKYGRAVDWWAFGVLIYOMLLQQSPFRGEDEDEIYDAILADEPLYPIHMPROSVSIL
                     QKILTREPDQRLGSGPTDAQEVMSQPFFRNIVWDDIYHKRVAPFLPQIKSAIDTSNF
                     DSEFTSVFPVLTPVQSVLSQAMQEEFRGFSYTADFD"
BASE COUNT     1447 a   1612 c   1479 g   1303 t
ORIGIN
```

```
Qy          413 gaggcacattggggcagc 431
                ||||||||||||||||||
```

Query Match   1.98;   Score 19;   DB 13;   Length 5841;
Best Local Similarity   100.0%;   Pred. No. 30;
Matches   19;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Db      464 GAGGGCACATTGGCGAGC 446

RESULT  42
LOCUS       CELC32B5     42545 bp      DNA              INV       05-DEC-1996
DEFINITION  Caenorhabditis elegans cosmid C32B5.
ACCESSION   U80843
VERSION     U80843.1  GI:1707218
KEYWORDS
SOURCE      Caenorhabditis elegans strain=Bristol N2.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 42545)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
REFERENCE   2  (bases 1 to 42545)
AUTHORS     Sheet,P.
TITLE       The sequence of C. elegans cosmid C32B5
JOURNAL     Unpublished (1996)
REFERENCE   3  (bases 1 to 42545)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (02-DEC-1996)
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO  63110, USA, and
            Sanger Centre, Hinxton Hall,
            Cambridge CB10 1RQ, England
            e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

            NOTICE: This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone.

            NEIGHBORING COSMID INFORMATION:

            The 5' cosmid is R07C3, 200 bp overlap;3' cosmid is F45D11. Actual
            start of this cosmid is at base position 1 of CELC32B5; actual end
            is at 42545 of CELC32B5

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1..42545
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="C32B5"
     gene            complement(538..2159)
                     /gene="C32B5.10"
     CDS             complement(join(538..670,762..1790,2089..2159))
                     /gene="C32B5.10"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37966.1"
                     /db_xref="GI:1707228"
                     /translation="MAYSLFITYPSLKCVIKHMEVNSRIQLSORCPSIRAEKSIPLK
                     IGHLELDQCRVKINNTDYSLAVVCNFSSPYDLDEVGSMDLGEVIDEQRDVVDNRSL
                     EEGEKRILREYELRMLTDDWEKKNALIGEALKKELKMTIEKSELVNRKCYVGIIKSL
                     FLSRRVGVPAPIWDIDFVNKNKLITFRDEKLSYNKLHQAFKYLIFKLFSGGRDAIQV
                     KKLSIENKGILRVPASLNLKVTHLEILGSGVDKIAISFQPLLEGSSLQSIKLNNDFNI
                     ENPILQTSPLLIFKNPENELFRARSAFTNVRVELEKSTIFPREALDLIANWLRHGKEV
                     GTIWNFSLTEVTALVCFKDIRKLPGAKYAGHRKTSSQLGRSVLIPINAFSEIHIFDEKE
                     EAQDGNDWSLKITVRRKL"
     gene            complement(3176..3625)
                     /gene="C32B5.11"
     CDS             complement(join(3176..3446,3500..3545,3589..3625))
                     /gene="C32B5.11"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37965.1"
                     /db_xref="GI:1707227"
                     /translation="MVIEMLGVVSVHFVNGDPAPCYSRHWNRNIELLDDKAESFWSK
                     FRVLLDPFKSIPFKSLELNGSISIDDNVIQSSELLIHNGRLDENIDLEFKLIKLANQQ
                     VELTTVYGSPNMVEL"
     gene            complement(4928..8090)
                     /gene="C32B5.12"
     CDS             complement(join(4928..5078,5462..6453,7974..8090))
                     /gene="C32B5.12"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37967.1"
                     /db_xref="GI:1707229"
                     /translation="MYSLSTLFITLFPKNFLSKYFIASWFLSFLKSSCFGLAIPPVA
                     YCPSLRAAEKSVPLHIKHLRITPHTIQINKTRLQLAIGQKGPDNEPTSYDIDQHGEKD
                     LGKIRPASDDEILDGRRHELTEEEKAANRPFEVAAIRENFRLRAFMADMVETDVEPT
                     EVRFPNHLGQLVRNSNQELESVIPSANWHVTPQARLGIFEPLSKGKGMRKLQERLI
                     ERVAYRKLHEIMKYLAEKLFGGKHLRIOKLEIGSNPGVLRLPNNMTIFTRNLELFDE
                     DHEYLLKSIRPLMNLKSLPLASLELHGPLPLESDVIETSKKLLLSEILDVNFDVERQE
                     IKKFKNQEVVLKFDEARAHFYESFEVPAGSWEAPGQGATEPPITFNIPMKYSSSKNEVS
                     IHCELDPNLRFHWVFSMTVQKPEIIDY"
     gene            8026..10315
                     /gene="C32B5.9"
     CDS             join(8026..8540,9362..10138,10195..10315)
                     /gene="C32B5.9"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37964.1"
                     /db_xref="GI:1707226"
                     /translation="MKYLLKKFFGGRVIVKVDKLIFCRIVLRIPASLKFQVHNLQLF
                     SSDNNQLPEAIKPILDFSTPLSSITLFYGYKINVEDPIIRSTETLYFDGYFLLLDNDL
                     ENLNKLRHKRVHLSLKPSSRWQNVPRLIDWIEVGREAGFYYSLEMNEEIRARELKT
                     VMINHKERITERFHLSQQYPSLLSTERSVPLRLDSLKLGGNSDSVNSIRCELAVYRHG
                     GRLSKKFKLDVTEDGRVDMNIVVVEGLGEIIVDLRGNREINTADVMKIFGGRGLIIVG
                     TLSIGCRGILRIPISRLKFEIRDLELGSSEDNIKLFEEIKPLLHLTSPPLNFPSLFNSYN
                     LRYEDPVVQSTGVLIFKNISSFDNNELENLNELRHKRAHLSFDRYFELQIIVGLIGNW
                     IKFGRIGTYSVELQAKNKGSEVLETIKRNHEERIIDRPNEDDNIVLLRMNNLLALY
                     ISYEFYEKYCLMHFRVQPRY"
     gene            10652..11939
                     /gene="C32B5.8"
     CDS             join(10652..10716,11060..11572,11628..11756,11807..11939)
                     /gene="C32B5.8"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37957.1"
                     /db_xref="GI:1707219"
                     /translation="MEKPKPLSYASTVIQYSDPNLRFRLSQQCSSLGSTDKSVPLHLE
```

```
RESULT 43
LOCUS       AC007191    43390 bp    DNA           PRI    02-APR-1999
DEFINITION  Homo sapiens chromosome 19, cosmid R34072, complete sequence.
ACCESSION   AC007191
```

```
Qy    217 gcttccaacttccagcag 235
          ||||||||||||||||||
Db   5471 GCTTCCAACTTCCAGCAG 5489

Query Match     1.9%;  Score 19;  DB 6;  Length 42545;
Best Local Similarity   100.0%;  Pred. No. 30;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
ACCESSION   AC007191
VERSION     AC007191.1  GI:4558643
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43390)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K.J., Gordon,L., Dias,J., Ramirez,M.,
            Stilwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A.,
            Garnes,J., Danganan,L., Erler,A., Christensen,M., Georgescu,A.,
            Avila,J., Liu,S., Attix,C., Andreise,T., Trankheim,M.,
            Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R.,
            Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Sanders,C., Ow,D.,
            Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S and Carrano,A.V.
  TITLE     Sequence analysis of a 1.9 Mb region in 19q13.2 between APOE and
            D19S412
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 43390)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from q centromere to telomere. Cosmid
            R34072 overlaps cosmid R28204 (AC006132) to the left from bases 1
            to 2,155 of this accession, and overlaps cosmid R26955 to the right
            from bases 37,677 to 43,390. Additional chromosome 19 map and
            sequence information may be obtained at:
            http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1..43390
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="R34072"
                     /chromosome="19"
                     /map="19q13.2 between APOE and D19S412"
                     /cell_line="SHL2-B"
                     /clone_lib="LL19NC03 R chromosome 19-specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from human-hamster hybrid 5HL2-B, which
                     carries chromosome 19 as its only human chromosome."
     repeat_region   35..306
     repeat_region   35..306
                     /rpt_family="AluSg"
     repeat_region   307..474
                     /rpt_family="AluJo/FRAM"
     repeat_region   477..787
                     /rpt_family="AluSx"
     repeat_region   complement(835..976)
                     /rpt_family="MIR"
     repeat_region   1142..1443
                     /rpt_family="AluSq"
     mRNA            join(<1533..1574,1658..2359)
                     /gene="GIPR"
                     /product="Human GIP receptor (GIPR) mRNA, partial cds."
     gene            <1533..2359
                     /gene="GIPR"
                     /note="gastric inhibitory polypeptide receptor"
     CDS             join(<1533..1574,1658..1864)
                     /gene="GIPR"
                     /note="GASTRIC INHIBITORY POLYPEPTIDE RECEPTOR PRECURSOR;
                     GIP-R; GLUCOSE-DEPENDENT INSULINOTROPIC POLYPEPTIDE
                     RECEPTOR"
                     /codon_start=1
                     /product="GIPR_HUMAN [AA 385- 466]"
                     /protein_id="AAD22674.1"
                     /db_xref="GI:4558645"
                     /translation="GFLVSVLVCFINKEVQSEIRRGWHHCRLRRSLGEEQRQLPERAF
                     RALPDSSGPGEVPTSRGLSSGTLPGPGNEASRELESYC"
     repeat_region   complement(1590..1709)
                     /rpt_family=(CGG)n"
     repeat_region   complement(1886..2236)
```

```
                     SLKLSGNSISVNAIEYELAIYIRHGGARPKFKVDVTEKGRVDMNIVVEEDPEEILVDL
                     RGNQERTIANVVTELANQRIENIEKMIQRRRNYFLVLTVGRVREVMIYERKLHEAVK
                     YLVEKLLGGRGIIKVDTLSIRSRGIIRIPSRLKRSSLDFVHHMHLHGKEIGSYYTFVI
                     TEQNINEAMSRIREEYEGIAWKTGEITINMKADSQLVITRIQQGDAWILGMKVKSRVV
                     LEY"
     gene            12939..14079
                     /gene="C32B5.16"
     CDS             join(12939..13009,13064..13903,13953..14079)
                     /gene="C32B5.16"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37971.1"
                     /db_xref="GI:1707233"
                     /translation="MSYSKPMSYPVIKCLLEHLDPNLRFKIALSCPLIKNLEKLVHLH
                     IKYLKLTDHSITINKDTYQFTVHNTKNGFLHDIDERDEDLDLSPIHEDPNEIIDVRKN
                     IRWYFPRAANVSSENTIOLKIIGDKFEILRYRRKMHEALKLYLIGKLFEGKTYKVNHLTIK
                     CGNVTRILSTVKFRVKILEVLDENTYDELDTIRSIIQTSSLPLNVETFRRLQAPVSRH
                     ELIRTAGMEIRCSIDVHTLATLNHRVHAPYFYEWRSNCFVDFLEIWMEYGKEIGSY
                     YTFNIFEHRIDKAMKRIRKMYKRSVEAGSARRAGQLTLNMNNSSQLVISKVKQGDAWI
                     LGVEVKDRWVL"
     gene            complement(15330..15886)
                     /gene="C32B5.13"
     CDS             complement(join(15330..15493,15542..15723,15780..15886))
                     /gene="C32B5.13"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37968.1"
                     /db_xref="GI:1707230"
                     /translation="MYAKANNRTVLSFSEQOIIDCGNFTSPCQENILSHEFIKKNGVV
                     TEADYPVVGKENEKCKYDENKIKLMPTNMLLVGNLPETLLKLFIKEHGPGYFRMKAPP
                     SFPNYKTIGYSPTQDEECGKATDARSLTIVGYGIEGGQNYWIVKGSFGT"
     gene            16984..18023
                     /gene="C32B5.7"
     CDS             join(16984..17051,17217..17561,17627..17802,17860..18023)
                     /gene="C32B5.7"
                     /note="strong similarity to the peptidase family C1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37963.1"
                     /db_xref="GI:1707225"
                     /translation="MKKSKPSSYQSLKTVLKHVDFNLREKPFDEKPANLRNTRNYIGLFL
                     LTALFYIIGALVQQRTQNYKNAKKPFLDWRDEGVVGPVKDQOGNCNASYAFAAISAIE
                     SMYAIANGQLLSFSEQOIIDCLGGCCAIESDPMMAMTYLERKGIETYTDYPFVGKNEK
                     CEYDSKKAYLILDDYDMSDESLALVFIDERGPGLFTMNTPPSFFNYKSGIYNPTEEE
                     CKSTNEKRALTIVGDNKGQNYMIVKGSFGT"
     gene            complement(19648..20937)
                     /gene="C32B5.14"
     CDS             complement(join(19648..19762,19822..20814,20867..20937))
                     /gene="C32B5.14"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB37969.1"
                     /db_xref="GI:1707231"
                     /translation="MSNSKPVNYLTLKCLLKNMDSGMRLQLFVMFPSIQYLEKLFPLH
                     VKLTIKSDHITVNRTTFQLKICNKYNNHGGYKPFDFDQYGRLDRGEIEQDPDESIID
                     VRDGFLSKKGIPECEMETARLVHNILTLQKSQRYSRIWESHGTILKKLSYVPENNFI
                     RLKIGKRVEVLEYQRKIHEAMKYLLGRLFGGRSLEANQFSIGCDTVLRVPSTLKFRIE
                     NLYPSFKIANTLDYVNQIVDNSSLPLSLKYSFENHIYHIHPHSLVRTVKMLKLEVEM
                     VPDIYISGIVSNLQMVDEKRAHIVFLGDCTSSNFLKITLAHWILEFHRDIGTYHTYQLSE"
```

```
repeat_region    complement(2357..2492)
                 /rpt_family="LINE2"
repeat_region    complement(2493..2771)
                 /rpt_family="AluJb"
repeat_region    complement(2786..3192)
                 /rpt_family="AluSx"
repeat_region    3223..3544
                 /rpt_family="LINE2"
misc_feature     3223..3544
                 /note="DDS similarity to overlapping ESTs:
                 (3223..3435) AA345686 EST51736 Gall bladder I Homo sapiens
                 cDNA 5' end; Score: 411 Identity: 211/214 (98%).
                 (3544..3253) AA987372 or92n03.s1 NCI_CGAP_Lu5 Homo sapiens
                 cDNA clone IMAGE:1603349 3'; Score: 575 Identity: 291/291
                 (100%).
                 (3544..3266) A483616 ne40e08.s1 NCI_CGAP_Co3 Homo sapiens
                 cDNA clone IMAGE:899846; Score: 550 Identity: 277/279
                 (99%)."
repeat_region    4318..4431
                 /rpt_family="MIR"
repeat_region    complement(4440..4738)
                 /rpt_family="AluSx"
repeat_region    complement(4916..5215)
                 /rpt_family="AluSc"
repeat_region    5362..5603
                 /rpt_family="AluSx"
repeat_region    complement(5766..5995)
                 /rpt_family="AluJb"
repeat_region    complement(6052..6349)
                 /rpt_family="MIR"
repeat_region    complement(6428..6469)
                 /rpt_family="AluSx"
repeat_region    complement(6656..6942)
                 /rpt_family="Alu"
repeat_region    complement(6943..7077)
                 /rpt_family="AluSq/x"
misc_feature     complement(7381..7656)
                 /note="DDS similarity to overlapping ESTs:
                 (7656..7381) AA983186 oq52h06.s1 NCI_CGAP_Kids Homo
                 sapiens cDNA clone IMAGE:1590011 3' similar to
                 SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2
                 ; (275..1) 98% identity.
                 (7656..7387) AI004535 ot63e01.s1 Soares_testis_NHT Homo
                 sapiens cDNA clone IMAGE:1621464 3' similar to
                 SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM
                 D2; (270..1) 99% identity.
                 (7656..7388) AA644001 nml9e12.s1 NCI_CGAP_Co10 Homo
                 sapiens cDNA clone IMAGE:1060654 similar to SW:SMD2_HUMAN
                 P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 ; (269..1)
                 99% identity.
                 (7656..7382) AI092009 oz97a08.x1
                 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
                 IMAGE:1683254 3' similar to SW:SMD2_HUMAN P43330 SMALL
                 NUCLEAR RIBONUCLEOPROTEIN SM D2 ; (275..1) 98% identity.
                 (7656..7391) W58351 zd19h08.r1 Soares fetal heart NbHH19W
                 Homo sapiens cDNA clone 341151 5' similar to SW:SMD2_HUMAN
                 P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 ; (212..478)
                 98% identity.
                 and others."
mRNA             complement(join(7482..7656,8316..8495,11816..11844))
                 and others.
                 /function="MRNA PROCESSING"
                 /product="Human SnRNP core protein Sm D2 mRNA, complete
                 cds"
CDS              complement(join(7381..7656,8316..8495,11816..11817))
                 and others.
                 /function="MRNA SPLICING"
                 /note="SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2; SNRNP CORE
                 PROTEIN D2; SM-D2"
                 /codon_start=1
                 /product="SMD2_HUMAN"
                 /protein_id="AAD22673.1"
                 /db_xref="GI:4558644"
                 /translation="MSLLNKPKSSMTPEELQKREEEEFNTGPLSVLTQSVKNNTQVLI
                 NCRNNKKLLGRVKAFPDRHCNMVLENVKEMWTEVPKSGKGKKKKSKPVNKDRYISKMFLR
                 GDSVIVVLRNPLIAGK"
```

```
repeat_region    7803..7912
                 /rpt_family="MIR"
repeat_region    complement(7923..8211)
                 /rpt_family="AluSx"
misc_feature     complement(8316..8497)
                 /note="DDS similarity to overlapping ESTs:
                 (8497..8316) AA983186 oq52h06.s1 NCI_CGAP_Kids Homo
                 sapiens cDNA clone IMAGE:1590011 3' similar to
                 SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM
                 D2; (457..276) 99% identity.
                 (8491..8316) AI004535 ot63e01.s1 Soares_testis_NHT Homo
                 sapiens cDNA clone IMAGE:1621464 3' similar to
                 SW:SMD2_HUMAN P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM
                 D2; (446..271) 99% identity.
                 (8497..8316) AA644001 nml9e12.s1 NCI_CGAP_Co10 Homo
                 sapiens cDNA clone IMAGE:1060654 similar to SW:SMD2_HUMAN
                 P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 ; (451..270)
                 99% identity.
                 (8495..8316) AI092009 oz97a08.x1
                 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
                 IMAGE:1683254 3' similar to SW:SMD2_HUMAN P43330 SMALL
                 NUCLEAR RIBONUCLEOPROTEIN SM D2 ; (455..276) 100%
                 identity.
                 (8495..8316) W58351 zd19h08.r1 Soares fetal heart NbHH19W
                 Homo sapiens cDNA clone 341151 5' similar to SW:SMD2_HUMAN
                 P43330 SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 ; (33..211)
                 98% identity.
                 and others."
repeat_region    complement(8666..8961)
                 /rpt_family="AluSc"
repeat_region    8962..9315
                 /rpt_family="AluSx"
repeat_region    9319..9446
                 /rpt_family="LINE2"
repeat_region    9447..9477
                 /rpt_family="AluSx"
repeat_region    9487..9661
                 /rpt_family="(CAAAA)n"
repeat_region    9681..9975
                 /rpt_family="AluSx"
repeat_region    9979..10278
                 /rpt_family="AluSx"
repeat_region    10279..10433
                 /rpt_family="AluY"
repeat_region    10479..10780
                 /rpt_family="AluSg/x"
repeat_region    10820..10903
                 /rpt_family="AluSx"
repeat_region    10903..11059
                 /rpt_family="LINE2"
repeat_region    complement(11059..11360)
                 /rpt_family="AluSx"
repeat_region    complement(11467..11599)
                 /rpt_family="FLAM_C"
```

```
RESULT 44
AP000333
LOCUS       AP000333    47404 bp    DNA    PRI    20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone:B353B2, complete sequence.
ACCESSION   AP000333
VERSION     AP000333.1   GI:4835702
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:B353B2.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Qy      7 ccacaagattcaaacccag 25
          |||||||||||||||||||
Db  29185 CCACAAGATTCAAACCCAG 29203

Query Match        1.9%;  Score 19;  DB 86;  Length 43390;
Best Local Similarity  100.0%;  Pred. No. 30;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

REFERENCE   1 (bases 1 to 47404)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE   Homo sapiens 47,404bp genomic DNA of 21q22.1
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2 (bases 1 to 47404)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ, 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT   The sequence is a part of the data (ACCESSION NO. AP000165 -
            AP000173).
            The sequencing project is supported by Japan Science Technology
            Corporation (JST) and The Institute of Physical and Chemical
            Research (RIKEN).
FEATURES   Location/Qualifiers
     source   1..47404
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="B353B2"
                     /map="21q22.1"
BASE COUNT   13800 a   9563 c 10009 g 14032 t
ORIGIN

Qy   641 ccagtgccaggaggtcctg 659
               ||||||||||||||||||||
Db   8162 CCAGTGCCAGGAGGTCCTG 8180

RESULT 45
AC091020/c
   Query Match   1.9%; Score 19; DB 91; Length 47404;
   Best Local Similarity   100.0%; Pred. No. 30;
   Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

LOCUS   AC091020   61435 bp   DNA   HTG   24-MAR-2001
DEFINITION Homo sapiens chromosome 18 clone RP11-701C7 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC091020
VERSION   AC091020.1 GI:13443176
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 61435)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE   Homo sapiens chromosome 18, clone RP11-701C7
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 61435)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,M., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,M. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------- Genome Center --------------------
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu

            -- Project Information
            Center project name: L12249
            Center clone name: 701_C_7
            ----
            * NOTE: This record contains 75 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            *   1   682: contig of 682 bp in length
            *   683   782: gap of 100 bp
            *   782   1495: contig of 713 bp in length
            *   1496   1595: gap of 100 bp
            *   1595   2335: contig of 740 bp in length
            *   2336   2435: gap of 100 bp
            *   2436   3147: contig of 712 bp in length
            *   3148   3247: gap of 100 bp
            *   3247   3964: contig of 717 bp in length
            *   3965   4064: gap of 100 bp
            *   4065   4755: contig of 691 bp in length
            *   4756   4855: gap of 100 bp
            *   4856   5569: contig of 714 bp in length
            *   5570   5669: gap of 100 bp
            *   5669   6349: contig of 680 bp in length
            *   6350   6449: gap of 100 bp
            *   6450   7167: contig of 718 bp in length
            *   7168   7267: gap of 100 bp
            *   7268   7989: contig of 722 bp in length
            *   7990   8089: gap of 100 bp
            *   8090   8828: contig of 739 bp in length
            *   8829   8928: gap of 100 bp
            *   8929   9638: contig of 710 bp in length
            *   9639   9738: gap of 100 bp
            *   9738   10460: contig of 722 bp in length
            *   10461   10560: gap of 100 bp
            *   10560   11296: contig of 736 bp in length
            *   11297   11396: gap of 100 bp
            *   11397   12093: contig of 697 bp in length
            *   12094   12193: gap of 100 bp
            *   12193   12905: contig of 712 bp in length
            *   12905   13005: gap of 100 bp
            *   13005   13725: contig of 720 bp in length
            *   13725   13825: gap of 100 bp
            *   13826   14540: contig of 715 bp in length
            *   14541   14640: gap of 100 bp
            *   14641   15287: contig of 647 bp in length
            *   15288   15387: gap of 100 bp
            *   15387   16101: contig of 714 bp in length
            *   16101   16201: gap of 100 bp
            *   16202   16876: contig of 675 bp in length

```
* 16877  16976: gap of 100 bp
* 16977  17740: contig of 764 bp in length
* 17741  17840: gap of 100 bp
* 17841  18522: contig of 682 bp in length
* 18523  18622: gap of 100 bp
* 18623  19353: contig of 731 bp in length
* 19354  19453: gap of 100 bp
* 19454  20186: contig of 733 bp in length
* 20187  20286: gap of 100 bp
* 20287  21003: contig of 717 bp in length
* 21004  21103: gap of 100 bp
* 21104  21821: contig of 718 bp in length
* 21822  21921: gap of 100 bp
* 21922  22634: contig of 713 bp in length
* 22635  22734: gap of 100 bp
* 22735  23430: contig of 696 bp in length
* 23431  23530: gap of 100 bp
* 23531  24226: contig of 696 bp in length
* 24227  24326: gap of 100 bp
* 24327  25057: contig of 731 bp in length
* 25058  25157: gap of 100 bp
* 25158  25918: contig of 761 bp in length
* 25919  26018: gap of 100 bp
* 26019  26753: contig of 735 bp in length
* 26754  26853: gap of 100 bp
* 26854  27578: contig of 725 bp in length
* 27579  27678: gap of 100 bp
* 27679  28378: contig of 700 bp in length
* 28379  28478: gap of 100 bp
* 28479  29202: contig of 724 bp in length
* 29203  29302: gap of 100 bp
* 29303  30065: contig of 763 bp in length
* 30066  30165: gap of 100 bp
* 30166  30876: contig of 711 bp in length
* 30877  30976: gap of 100 bp
* 30977  31677: contig of 701 bp in length
* 31678  31777: gap of 100 bp
* 31778  32495: contig of 718 bp in length
* 32496  32595: gap of 100 bp
* 32596  33291: contig of 696 bp in length
* 33292  33391: gap of 100 bp
* 33392  34102: contig of 711 bp in length
* 34103  34202: gap of 100 bp
* 34203  34972: contig of 770 bp in length
* 34973  35072: gap of 100 bp
* 35073  35795: contig of 723 bp in length
* 35796  35895: gap of 100 bp
* 35896  36671: contig of 776 bp in length
* 36672  36771: gap of 100 bp
* 36772  37483: contig of 712 bp in length
* 37484  37583: gap of 100 bp
* 37584  38290: contig of 707 bp in length
* 38291  38390: gap of 100 bp
* 38391  39091: contig of 701 bp in length
* 39092  39191: gap of 100 bp
* 39192  39910: contig of 719 bp in length
* 39911  40010: gap of 100 bp
* 40011  40740: contig of 730 bp in length
* 40741  40840: gap of 100 bp
* 40841  41571: contig of 731 bp in length
* 41572  41671: gap of 100 bp
* 41672  42431: contig of 760 bp in length
* 42432  42531: gap of 100 bp
* 42532  43249: contig of 718 bp in length
* 43250  43349: gap of 100 bp
* 43350  44088: contig of 739 bp in length
* 44089  44188: gap of 100 bp
* 44189  44950: contig of 762 bp in length
* 44951  45050: gap of 100 bp
* 45051  45813: contig of 763 bp in length
* 45814  45913: gap of 100 bp
* 45914  46609: contig of 696 bp in length
* 46610  46709: gap of 100 bp

* 46710  47423: contig of 714 bp in length
* 47424  47523: gap of 100 bp
* 47524  48208: contig of 685 bp in length
* 48209  48308: gap of 100 bp
* 48309  49010: contig of 702 bp in length
* 49011  49110: gap of 100 bp
* 49111  49889: contig of 779 bp in length
* 49890  49989: gap of 100 bp
* 49990  50577: contig of 588 bp in length
* 50578  50677: gap of 100 bp
* 50678  51404: contig of 727 bp in length
* 51405  51504: gap of 100 bp
* 51505  52236: contig of 732 bp in length
* 52237  52336: gap of 100 bp
* 52337  53095: contig of 759 bp in length
* 53096  53195: gap of 100 bp
* 53196  53943: contig of 748 bp in length
* 53944  54043: gap of 100 bp
* 54044  54802: contig of 759 bp in length
* 54803  54902: gap of 100 bp
* 54903  55630: contig of 728 bp in length
* 55631  55730: gap of 100 bp
* 55731  56410: contig of 680 bp in length
* 56411  56510: gap of 100 bp
```

```
Query Match        1.9%;   Score 19;  DB 78;  Length 61435;
Best Local Similarity   100.0%;  Pred. No. 30;
Matches  19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    211 gctgcagcttcccaacttc 229
          |||||||||||||||||||
Db  55825 GCTGCAGCTTCCCAACTTC 55807
```

Search completed: October 28, 2001, 22:55:01
Job time: 4014 sec

OM nucleic - nucleic search, using sw model

```
                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:             October 28, 2001, 21:51:47 ; Search time 102.68 Seconds
                                                 (without alignments)
                                                 6121.246 Million cell updates/sec

Title:              SEQLINS_COPY_1140_2140
Perfect score:      1001
Sequence:           1 tcgcagccacaagattcaaa.........aagcctcagcgtgggatg 1001

Scoring table:      OLIGO_NUC
                    Gapop 60.0 , Gapext 60.0
Searched:           730101 seqs, 313950809 residues
Word size :         8

Total number of hits satisfying chosen parameters:      731206

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  N_Geneseq_0601:*
      1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
      2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
      3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
      4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
      5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
      6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
      7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
      8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
      9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
     10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
     11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
     12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
     13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
     14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
     15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
     16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
     17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
     18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
     19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
     20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
     21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
     22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	880	87.9	2546	21	AAC76445	Human ORFX ORF2000
2	880	87.9	2958	21	AAA58453	Human prostate can
3	727	72.6	2478	21	AAA52810	Human sulphatase G
4	124	12.4	26664	21	AAA60207	Human prostate can
5	24	2.4	24	21	AAA60374	Human prostate can
6	20	2.0	457	21	AAC57035	Pinus radiata tran
7	19	1.9	19	21	AAA60229	Human prostate can
8	19	1.9	19	21	AAA60230	Human HPC2 cDNA se
9	18	1.8	18	21	AAA60225	Human HPC2 cDNA se
10	18	1.8	18	21	AAA60226	Human HPC2 cDNA se
11	18	1.8	18	21	AAA60227	Human HPC2 cDNA se

```
c 12            18  21  AAA60228  Human HPC2 cDNA se
c 13            18  21  AAA60237  Human HPC2 cDNA se
  14            18  21  AAA60238  Human HPC2 cDNA se
  15            18  21  AAA60240  Human HPC2 cDNA se
  16            18  20  AAX18060  Coding sequence fo
  17            18  18  AAF67773  H. Pylori cytoplas
  18            18  18  AAT77453  H. Pylori cytoplas
  19            18  18  AAC44859  Arabidopsis thalia
c 20            18  18  AAT67992  Human ORFX ORF2745
c 21            18  21  AAC77190  Human ORFX ORF2745
c 22            18  19  AAX08683  Novel nucleotide s
c 23            18  20  AAV17351  Coding sequence fo
c 24            18  21  AAC09473  H. helicase
c 25            17  21  AAA60239  Human HPC2 cDNA se
c 26            17  21  AAA60241  Human HPC2 cDNA se
c 27            17  21  AAA60242  Human HPC2 cDNA se
c 28            17  21  AAA32079  Plant microsatelli
c 29            17  21  AAC32768  Arabidopsis thalia
c 30            17  21  AAA32050  Plant microsatelli
c 31            17  21  AAA42643  Plant microsatelli
c 32            17  16  AAT24629  Human 5' EST isola
c 33            17  21  AAT24629  Human gene signatu
c 34            17  21  AAC07468  Human secreted pro
c 35            17  22  AAF64559  Novel human polynu
  36            17  21  AAA32069  Novel human polynu
  37            17  21  AAA32052  Plant microsatelli
  38            17  21  AAC45096  Arabidopsis thalia
c 39            17  19  AAC45096  Arabidopsis thalia
c 40            17  20  AAV29360  Calcium ion channe
c 41            17  22  AAF64735  Novel human polynu
c 42            17  22  AAX24979  Clone G3156 encodi
  43            17  21  AAX291823 Streptococcus pneu
  44            17  15  AAQ67866  H6/MAGE-1 encoding ge
  45            17  20  AA208442  H6/MAGE-1 expressi
```

ALIGNMENTS

```
RESULT   1
AAC76445
ID   AAC76445 standard; cDNA; 2546 BP.
XX
AC   AAC76445;
XX
DT   08-FEB-2001   (first entry)
XX
DE   Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX
Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnerary; antipsoriatic; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
Immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   cholesterol ester storage; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   severe combined immunodeficiency; systemic lupus erythematosus; infection;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999; 99US-0127607.
```

```
PR   02-APR-1999; 99US-0127636.
PR   05-APR-1999; 99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach M;
XX
DR   WPI; 2000-603362/57.
DR   P-PSDB; AAB42236.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3179-3180; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in open reading frames
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;
```

```
Query Match             87.9%;
Best Local Similarity   99.9%;
Matches 1000;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

                        Score 880;  DB 21;  Length 2546;

Qy      1 tcgcagccacaagaattcaaacccagctcaatccaccccggacactctcccctgt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 tcgcagccacaagaattcaaacccagctcaatccaccccggacactctcccctgt 797

Qy     61 caccagtttccgctgtaagaagaagaggagggcccacccctcagtgtgccaggtgta 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 caccagtttccgctgtaagaagaagaggagggcccacccctcagtgtgccaggtgta 857

Qy    121 atgcctcctcaagtacaagctccgtcccagaaggggatgccattattac 180
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 atgcctcctcaagtaccagctccgtcccagaaggggatgccattattac 917

Qy    181 ttgtcaatcctgaggaattcatagttgagcgctgcagctcccaacttccagcgagcgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 ttgtcaatcctgaggaattcatagttgagcgctgcagctcccaacttccagcgagcgt 977

Qy    241 gcaggagtacagcaggaggagtgcgcaggaacgacgcccagcacagaagaagaagtcagta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 gcaggagtacagcaggaggagtgcgcaggaacgacgcccagcacagaagaagaagtcagta 1037

Qy    301 cccagaaatcatcttcctggaacaggtgtctgccatcccgatgaagttcgaaaatgtcag 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 cccagaaatcatcttcctggaacaggtgtctgccatcccgatgaagttcgaaaatgtcag 1097

Qy    361 tgccacacttgtcaaacataagccccgacacgtctctgtactgtgttgaggggcac 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 tgccacacttgtcaaacataagccccgacacgtctctgtactgtgttgaggggcac 1157

Qy    421 atttgggcagctgtgccgtcattacgagcagatccagatcagagtggacagagtcctggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 atttgggcagctgtgccgtcattacgagcagatccagatcagagtggacagagtcctggc 1217

Qy    481 tgctgtgtttgtgtgtccccaccgcacgcagatccacacacgggcttgcaagtatcttgc 540
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   1218 tgctgtgtttgtgtgtccccacct-gcacgcagatccacacacgggcttgcaagtatcttgc 1276

Qy    541 tgcagagagaacgcgccttggcatctttgggaaagcgagaaagccgcttcaccttgctg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1277 tgcagagagaacgcgccttggcatctttgggaaagcgagaaagccgcttcaccttgctg 1336

Qy    601 ccccaaccagctcaaagcctggctccagcagtaccacaaccagtccaggaggtcctgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1337 ccccaaccagctcaaagcctggctccagcagtaccacaaccagtccaggaggtcctgc 1396

Qy    661 accacatcagtatgatcctgccaaatgccttcaggagggctgagatctccagtcctg 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1397 accacatcagtatgatcctgccaaatgccttcaggagggctgagatctccagtcctg 1456

Qy    721 cagtggaaaagattgatcaagttcgctgttgccgaacacatgtgagagtttcagacct 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1457 cagtggaaaagattgatcaagttcgctgttgccgaacacatgtgagagtttcagacct 1516

Qy    781 gtctgtgtcggcacctgcaagcatgtggctgtcgctgttggcaacctctggctga 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1517 gtctgtgtcggcacctgcaagcatgtggctgtcgctgttggcaacctctggctgga 1576

Qy    841 aagtggtctattccgggacaccatgccctgccctgcgcaggctctggtccggatgggaaagatg 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1577 aagtggtctattccgggacaccatgccctgccctgcgcaggctctggtccggatgggaaagatg 1636

Qy    901 ccaccctcctgatacatggaagcacacctgaagatgtttggaggaggcagtgaaa 960
          |||||||||||||||||| |||||||||| ||||||||||||||||| ||||||||
Db   1637 ccaccctcctgatacatgaagcacacctggaagatgtttggaggaggaagcagtgaaa 1696

Qy    961 agacaacacagcacaacgtcccaaagccatcaagcgtggggatg 1001
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1697 agacaacacagcacaacgtcccaaagccatcaagcgtggggatg 1737
```

```
RESULT 2
AAA58453
ID   AAA58453 standard; cDNA; 2958 BP.
XX
AC   AAA58453;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..2531
FT                   /*tag= a
FT                   /product= "HPC2"
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
```

```
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Claim 3; Page 98-100; 157pp; English.
XX
CC   The present sequence is the coding sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 1'p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;
```

```
Query Match             87.98%;  Score 880;  DB 21;  Length 2958;
Best Local Similarity   99.9%;
Matches 1000;  Conservative    0;  Pred. No. 0;  Mismatches    0;  Indels    1;  Gaps    1;
```

```
Qy      1 tcgcagcacaagattcaaacctcaccaccggacatcctcccctgct 60
          ||||||| ||||||||||||||||| ||||||||||||||||||||
Db   1190 tcgcagccacaagattcaaacctcaacctcatccaccggacatcctcccctgct 1249

Qy     61 caccagtttccgctgtaagaagaggaggccccaccctcagtgtgcctgccatggttcaggtga 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 caccagtttccgctgtaagaagaggaggccccaccctcagtgtgccctgccatggttcaggtga 1309

Qy    121 atggcctcctcaagtacaagctccgtcccagggaggagtggcaggggatgtcagtta 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1310 atggcctcctcaagtacaagctccgtcccagggaggagtggcaggggatgtcattattac 1369

Qy    181 ttgcaatcctgaggaattcatagttgaggcgctgcagcttccaacttccagcaggcgt 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1370 ttgcaatcctgaggaattcatagttgaggcgctgcagcttccaacttccagcaggcgt 1429

Qy    241 gcaggagtacaggaggagtgcgcagacgacgcccagccccagcaagaaaagaagtcagta 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1430 gcaggagtacaggaggagtgcgcagacgacgcccagccccagcaagaaaagaagtcagta 1489

Qy    301 cccagaaatcatcctcttggaacaggtctgcaatccatcccagagaggcgt 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1490 cccagaaatcatcctcttggaacaggtctgcaatccatcccagagaggcgt 1549

Qy    361 tgccacacttgtcaacataagcccagacactctcgtactggactgtgttgaaggggac 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1550 tgccacacttgtcaacataagcccagacactctcgtactggactgtgttgaaggggac 1609

Qy    421 atttgggcagctgtcgtcattacggagaaccaaggggttcctgggcaccctggc 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1610 atttgggcagctgtcgtcattacggagaaccaaggggttcctgggcaccctggc 1669

Qy    481 tgctgtgtttgtcttcccactgggacgacgaagtggacaggtcctgggcaccctggc 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1670 tgctgtgtttgtcttcccacct gcacgcagaatcacacacgaagtggacaggtatcttgc 1728

Qy    541 tgcagagagaaacgcgccatcttgggaaagcgcttcaccctttgctggttggttg 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1729 tgcagagagaaacgcgccatcttggaaaagcgcttcaccctttgcgggtggttg 1788

Qy    601 cccccaaccacagctcaaagcctggctcagcgagcgtaccaacacagtgccaggaggtcctgc 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1789 cccccaaccacagctcaaagcctggctcagcgagcgtaccaacacagtgccaggaggtcctgc 1848

Qy    661 accacatcagtgatgattcctgccaaatgccttcaggaggggctgaatctccagtcctg 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1849 accacatcagtgatgattcctgccaaatgccttcaggaggggctgagatctccagtcctg 1908
```

```
Qy    721 cagtggaaagaattgatcagttcgctgttgcgaaacatgtgattttggaagaggtttcagacct 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1909 cagtggaaagaattgatcagttcgctgttgcgaaacatgtgattttggaagaggtttcagacct 1968

Qy    781 gtctggttgcgcactgcaagcacatgcgtttggctgtctgcgcggttgcacaacctctggctga 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1969 gtctggttgcgcactgcaagcacatgcgtttggctgtctgcgcggttgcacaacctctggctga 2028

Qy    841 aagtggttctattccgggacaccatgccctgcgaggctctggttccggatccgagaaagatg 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2029 aagtggttctattccgggacaccatgccctgcgaggctctggttccggatgtgggaaagatg 2088

Qy    901 ccaccctcctgatacatgaagccaccctggaagatggtttggaagagaggcagtcagtggaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2089 ccaccctcctgatacatgaagccaccctggaagatggtttggaagagaggcagtcagtggaaa 2148

Qy    961 agacacacagcacaacgtcccaagccatcagcgtggggatg 1001
          |||||||||||||||||||||||||||||||||||||||||
Db   2149 agacacacagcacaacgtcccaagccatcagcgtggggatg 2189
```

```
RESULT  3
AAA52810
ID   AAA52810 standard; cDNA; 2478 BP.
XX
AC   AAA52810;
XX
DT   20-SEP-2000 (first entry)
XX
DE   Human sulphatase G cDNA.
XX
KW   Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..2478
FT                   /*tag= a
FT                   /partial
FT                   /product= "hSG"
XX
PN   WO200034327-A1.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999; 99WO-AU01092.
XX
PR   09-DEC-1998; 98AU-0007624.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ, Litjens T, Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   P-PSDB; AAY99850.
XX
PT   Novel isolated DNA sequence which encodes human sulfatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency -
XX
PS   Claim 2; Page 29-30; 33pp; English.
XX
CC   The present sequence encodes human sulphatase G (hSG). hSG is
CC   not a member of the well-characterised CTPSR sulphatase family.
CC   It belongs to a family showing sequence similarity to a sulphatase
CC   from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC   contains 23 exons and is located at chromosome 17p11.2. The present
CC   sequence is clone lambda29.1 from a human testes cDNA library. It was
CC   isolated using human EST sequences with sequence similarity to the
CC   non-CTPSR family as a probe to screen the library. The cDNA
CC   subcloned and the DNA sequence of both strands was determined. The
CC   sequence may be used to treat a patient suffering from hSG deficiency
```

```
CC   by replacing, repairing, or compensating for a DNA sequence within that
CC   patient's genome.
XX
SQ   Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;

     Query Match          72.6%;  Score 727; DB 21; Length 2478;
     Best Local Similarity   99.6%;
     Matches  997; Conservative   0; Mismatches   3; Indels   1; Gaps   1;

Qy        1 tcgcagccacaagatccaaaaccagctcaacctcatccaccgagacatcttccccctgct   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1140 tcgcagccacaagatccaaaaccagctcaacctcatccaccgagacatcttccccctgct 1199

Qy       61 caccagtttccgctgtaagaaggaaggaggccccaccctcagtgtgcccatgttcagggtga  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1200 caccagtttccgctgtaagaaggaaggaggccccaccctcagtgtgccatgttcagggtga 1259

Qy      121 atgcctcctcaagtaccagctccgtcccgagagaggagtgggcaggaggatgccattac  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1260 atgcctcctcaagtaccagctccgtcccgagagaggagtgggcaggaggatgccattac 1319

Qy      181 ttgcaatcctgaggagattcattagttgaggcgctgcagcttcccaacttccagcagagcgt  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1320 ttgcaatcctgaggagattcattagttgaggcgctgcagcttcccaacttccagcagagcgt 1379

Qy      241 gcaggagtacaagaggagtcgcgacgagacgcccgagagggcaagaagaagtcagta  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1380 gcaggagtacaagagagtcgcgacgagacgcccgagagggcaagaagaagtcagta 1439

Qy      301 cccagaaatcatcttcctggaacagggtctgccatcccgatgaaaagattcgaaatgtcag  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1440 cccagaaatcatcttcctggaacagggtctgccatcccgatgaaaagattcgaaatgtcag 1499

Qy      361 tgctccacttgtcaaacataagcccccgacacagtctctgctactggactgtgtgaggcac  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1500 tgctccacttgtcaaacataagcccccgacacagtctctgctactggactgtgtgaggcac 1559

Qy      421 atttgggcagctgtgccgtcattacggagaccaagtgtccagctgccaccctggc  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1560 gtttgggcagctgtgccgtcattacggagaccaagtgtcctggccaccctggc 1619

Qy      481 tgctgtgttgtgtcccgcacgcaagaatcacacaggggcttgccaagtatcttgc  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1620 tactgtgtttgtgtcccacct-gcacgcaagatcacacaggggcttgccaagtatcttgc 1678

Qy      541 tgcagagagaacgcgcgcctggcatctcttggaaagcgcttcacccttgctggtggttg  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1679 tgcagagagaacgcgcgcctggcatctcttggaaagcgcttcacccttgctggtggttg 1738

Qy      601 cccccaaccagctcaaagctggctccagcaagtaccaacctgccaggaggtcctgc  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1739 cccccaaccagctcaaagctggctccagcaagtaccacaccagtgccaggaggtcctgc 1798

Qy      661 accaacatcagtatgattgattccctgccaaatgctcaggggctgagatctccagctcctg  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1799 accacaatcagtatgattgattccctgccaaatgcttcagggggctgagatctccagctcctg 1858

Qy      721 cagtggaaaagattgatcagttgcgctgttgaacatgtgttcaggaagaagttcagacct  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1859 cagtggaaaaagattgatcagttcccctgttgcgaacatgtgttttggaagagttcagacct 1918

Qy      781 gtctcgtgtcggcactgcaagactgcggctgcgctggtcgtcacactctgctgga  840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1919 gtctggtgtcggcactgcaagcgatgcgtttggctgtgctcgtcacactctgctgga 1978

Qy      841 aagtggtctattccgggacaccatgcctgttcgggtcgggtgggaaagatg  900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1979 aagtggtctattccgggacaccatgcctgttcgggtcgggtgggaaagatg 2038

Qy      901 ccaccctcctgatacatgaagcccaccctgaaagatggttggaaagaagagaagcagtcgaa  960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2039 ccaccctcctgatacatgaagcccaccctgaaagatggttggaaagaagagaagcagtcgaa 2098

Qy      961 agacacacaagcacacaacgtcccaaagccatcaagccgtggggatg  1001
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db     2099 agacacacaagcacacaacgtcccaaagccatcaagccgtggggatg 2139
```

```
RESULT   4
AAA60207
ID   AAA60207 standard; DNA; 26664 BP.
XX
AC   AAA60207;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 genomic sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
CDS             910..26039
FT                   /*tag= a
FT                   /product= "HPC2"
FT                   /note= "this sequence contains introns"
FT                   /transl_except= (pos:23892..23895,aa:Glu)
FT   exon            910..1154
FT                   /*tag= b
FT                   /number= 1
FT   exon            1736..1786
FT                   /*tag= c
FT                   /number= 2
FT   exon            1925..1995
FT                   /*tag= d
FT                   /number= 3
FT   exon            3025..3089
FT                   /*tag= e
FT                   /number= 4
FT   exon            4361..4418
FT                   /*tag= f
FT                   /number= 5
FT   exon            5582..5650
FT                   /*tag= g
FT                   /number= 6
FT   exon            7075..7194
FT                   /*tag= h
FT                   /number= 7
FT   exon            8186..8244
FT                   /*tag= i
FT                   /number= 8
FT   exon            12878..12936
FT                   /*tag= j
FT                   /number= 9
FT   exon            13032..13104
FT                   /*tag= k
FT                   /number= 10
FT   exon            13756..13868
FT                   /*tag= l
FT                   /number= 11
FT   exon            15283..15378
FT                   /*tag= m
FT                   /number= 12
FT   exon            16278..16416
FT                   /*tag= n
FT                   /number= 13
FT   exon            16498..16583
FT                   /*tag= o
FT                   /number= 14
FT   exon            18583..18701
FT                   /*tag= p
FT                   /number= 15
FT   exon            20349..20445
```

FT   exon   /*tag= q
FT   /number= 16
FT   22172..22310
FT   exon   /*tag= r
FT   /number= 17
FT   22879..22917
FT   exon   /*tag= s
FT   /number= 18
FT   23045..23154
FT   exon   /*tag= t
FT   /number= 19
FT   23795..23895
FT   exon   /*tag= u
FT   /number= 20
FT   23973..24093
FT   exon   /*tag= v
FT   /number= 21
FT   24354..24432
FT   exon   /*tag= w
FT   /number= 22
FT   25026..25170
FT   exon   /*tag= x
FT   /number= 23
FT   25812..26036
FT   exon   /*tag= y
FT   /number= 24
FT   26447..26452
FT   polyA_signal   /*tag= z
PN   XX
PA   WO20002786A-A1.
PD   XX
PF   18-MAY-2000.
XX
PR   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Claim 3;   Page 108-122; 157pp; English.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   P-PSDB; AAB07228.
DR   WPI; 2000-376481/32.
XX
PF   18-MAY-2000.
PD   XX
PR   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
SQ   Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;

CC   The present sequence is the genomic sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX

Query Match   12.4%; Score 124; DB 21; Length 26664;
Best Local Similarity   100.0%; Pred. No. 1.4e-52;
Matches 124; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   769 agtttcagaacctgtgtcggtggcggcactgcaagacatgcgtttggctgtgtggtgcaca 828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23971 agtttcagaacctgtgtcggtggcggcactgcaagacatgcgtttggctgtgtggtgcaca 24030

Qy   829 cctctggctggaaagtggtctattccgggggaccaccatgccctgcgaggctctggtccgga 888
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24031 cctctggctggaaagtggtctattccgggggaccaccatgccctgcgaggctctggtccgga 24090

Qy   889 tggg 892
             ||||
Db   24091 tggg 24094

RESULT   5
AC   AAA60374
ID   AAA60374 standard; cDNA; 24 BP.
XX
AC   AAA60374;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 variant 1641insG.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786A-A1.
XX
PF   05-NOV-1999;   99WO-US26055.
PD   18-MAY-2000.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 5;   Page 63; 157pp; English.
XX
CC   The present sequence is part of the coding sequence of a variant of the
CC   human prostate cancer predisposing gene HPC2, which is found on
CC   chromosome 17p. This variant has been designated 1641insG. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer, this allele being an example of this. The HPC2 gene and
CC   its protein can be used in peptide and gene therapy for cancer patients,
CC   as well as being useful as diagnostic tools (both for cancer sufferers
CC   and those with a predisposition to the disease) and in the production of
CC   cancer drugs. This sequence was isolated by mutation screening of the
CC   HPC2 gene in humans.
XX
SQ   Sequence 24 BP; 5 A; 9 C; 6 G; 4 T; 0 other;

Query Match   2.4%; Score 24; DB 21; Length 24;
Best Local Similarity   100.0%; Pred. No. 0.03;
Matches   24; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   491 gtgtcccacctggcacgcagatca 514
             ||||||||||||||||||||||||
Db   1 gtgtcccacctggcacgcagatca 24

RESULT   6
ID   AAC57035 standard; DNA; 457 BP.
XX
AC   AAC57035;
XX
DT   25-JAN-2001 (first entry)
XX
DE   Pinus radiata transcription factor DNA sequence #481.
XX

```
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Pinus radiata.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999; 99US-0266513.
PR   18-AUG-1999; 99US-0149485.
XX
DR   WPI; 2000-579369/54.
XX
PI   Wood M, McGrath A, Shenk MA, Glenn M;
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Pages 570-571; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 457 BP; 130 A; 94 C; 118 G; 115 T; 0 other;

Query Match             2.0%;  Score 20;  DB 21;  Length 457;
Best Local Similarity   100.0%;  Pred. No. 3;
Matches   20;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        944 agaggaagcagtggaaaga 963
              |||||||||||||||||||
Db        146 agaggaagcagtggaaaga 165

RESULT   7
AAA60229
ID   AAA60229 standard; DNA; 19 BP.
XX
AC   AAA60229;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 50.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
DR   WPI; 2000-376481/32.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 19 BP; 5 A; 3 C; 6 G; 5 T; 0 other;

Query Match             1.9%;  Score 19;  DB 21;  Length 19;
Best Local Similarity   100.0%;  Pred. No. 9.8;
Matches   19;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        763 tggaaagagtttcagacctg 781
              |||||||||||||||||||
Db          1 tggaaagagtttcagacctg 19

RESULT   8
AAA60230/c
ID   AAA60230 standard; DNA; 19 BP.
XX
AC   AAA60230;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 51.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26055.
XX
PR   06-NOV-1998; 98US-0107468.
XX
DR   WPI; 2000-376481/32.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Example 3; Page 55; 157pp; English.
```

```
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 19 BP; 5 A; 6 C; 3 G; 5 T; 0 other;

Qy   763 tggaagagttcaagacctg 781
             |||||||||||||||||||
Db    19 TGGAAGAGTTCAAGACCTG 1

Query Match          1.9%;   Score 19;   DB 21;   Length 19;
Best Local Similarity  100.0%;   Pred. No. 9.8;
Matches   19;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT   9
ID   AAA60225 standard; DNA; 18 BP.
XX
AC   AAA60225;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 46.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786 4-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 5 C; 3 G; 5 T; 0 other;


Qy   118 tgaatgcctctcaagta 135
             ||||||||||||||||||
Db     1 tgaatgcctctcaagta 18

Query Match          1.8%;   Score 18;   DB 21;   Length 18;
Best Local Similarity  100.0%;   Pred. No. 31;
Matches   18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT  10
ID   AAA60226/c
ID   AAA60226 standard; DNA; 18 BP.
XX
AC   AAA60226;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 47.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20002786 4-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 3 C; 5 G; 5 T; 0 other;


Qy   118 tgaatgcctctcaagta 135
             |||||||||||||||||
Db    18 TGAATGCCTCCTCAAGTA 1

Query Match          1.8%;   Score 18;   DB 21;   Length 18;
Best Local Similarity  100.0%;   Pred. No. 31;
Matches   18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT  11
ID   AAA60227
ID   AAA60227 standard; DNA; 18 BP.
XX
AC   AAA60227;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 48.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
```

```
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
XX
SQ   Sequence 18 BP; 3 A; 3 C; 7 G; 5 T; 0 other;

XX
XX
PS   Example 3; Page 55; 157pp; English.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX

Qy  397 gctactggactgtggtga 414
        ||||||||||||||||||
Db    1 gctactggactgtggtga 18

Query Match          1.8%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 12
AAA60228/c
ID   AAA60228 standard; DNA; 18 BP.
XX
AC   AAA60228;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 49.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
```

```
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 55; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 7 C; 3 G; 3 T; 0 other;

XX
XX
Qy  397 gctactggactgtggtga 414
        ||||||||||||||||||
Db   18 GCTACTGGACTGTGGTGA 1

Query Match          1.8%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


RESULT 13
AAA60237
ID   AAA60237 standard; DNA; 18 BP.
XX
AC   AAA60237;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 58.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;  99WO-US26055.
XX
PR   06-NOV-1998;  98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PS   Example 3; Page 56; 157pp; English.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 6 A; 7 C; 3 G; 2 T; 0 other;
```

RESULT 14
AAA60238
```
ID   AAA60238 standard; DNA; 18 BP.
XX
AC   AAA60238;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 59.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; HPC2;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 0 A; 3 C; 7 G; 8 T; 0 other;

Query Match      1.8%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches   18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   371 gtcaacataagcccccgac 388
         ||||||||||||||||||
Db     1 gtcaacataagcccgac 18
```

RESULT 15
AAA60240
```
ID   AAA60240 standard; DNA; 18 BP.
XX
AC   AAA60240;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 61.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 18 BP; 5 A; 4 C; 3 G; 6 T; 0 other;

Query Match      1.8%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches   18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   478 ggctgctgtgttgtgtc 495
         ||||||||||||||||||
Db     1 ggctgctgtgttgtgtc 18
```

RESULT 16
AAX18066/c
```
ID   AAX18066 standard; DNA; 177 BP.
XX
AC   AAX18066;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Coding sequence for human SI binding protein SNi45.
XX
KW   Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW   D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW   intestinal peptide-associated transporter; hypertension; diabetes;
KW   osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW   therapeutic agent delivery; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851325-A2.
XX
PD   19-NOV-1998.
XX
PF   15-MAY-1998;   98WO-US10088.

Query Match      1.8%;  Score 18;  DB 21;  Length 18;
Best Local Similarity  100.0%;  Pred. No. 31;
Matches   18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   671 tatgattcctgccaaatg 688
         ||||||||||||||||||
Db     1 tatgattcctgccaaatg 18
```

PR   15-MAY-1997; 97US-0046595.
XX
PA   (CYTO-) CYTOGEN CORP.
PA   (ELAN-) ELAN CORP PLC.
XX
PI   Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
PI   Omahony DJ, Patterson CA, Singleton J;
XX
DR   WPI; 1999-009568/01.
XX
PS   Claim 49; Page 56; 294pp; English.
XX
CC   New proteins that bind specifically to receptors in the
CC   gastro-intestinal tract and related nucleic acid - chimaeras and
CC   proteins (I) that bind specifically to at least one of the
CC   gastro-intestinal (GI) tract receptors human intestinal
CC   peptide-associated transporter (HPT1), hPEPT1, D2H and human
CC   sucrose-isomaltase complex (hSI). (I) provide active transport of
CC   therapeutic agents through human and animal GI tissue (into the blood)
CC   for in vivo delivery, particularly for treatment or prevention of
CC   hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC   migraine, or angina pectoris. Specifically they are used to deliver
CC   insulin or leuprolide, but many other suitable therapeutic agents are
CC   disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC   antigens. (I) may also provide targeting to the GI tract. Other uses of
CC   (I) are: (i) to determine the level of specified receptors in a sample
CC   (in a binding assay); and (ii) to screen for molecules that bind (I).
CC   Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC   and in immunoassays. The antibodies are used to locate, detect and
CC   measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC   etc., also for peptide purification and immobilisation.
XX
SQ   Sequence 177 BP; 31 A; 51 C; 53 G; 42 T; 0 other;

Qy   628 agcagtaccacaaccagt 645
         ||||||||||||||||||
Db   90 AGCAGTACCACAAACCAGT 73

RESULT 17
ID   AAT67773/c
AC   AAT67773;
XX
DT   29-JUL-1997 (first entry)
XX
DE   H. Pylori cytoplasmic protein ORF 24824087.aa.
XX
KW   Vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   cytoplasmic; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..531
FT   /*tag= a
FT   /transl_except= (pos: 460..462, aa: Xaa)
FT   /transl_except= (pos: 520..522, aa: Xaa)
FT   /transl_except= (pos: 526..528, aa: Xaa)
FT   /note= "Xaa = Unknown"

XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996; 96WO-US09122.
XX
PR   01-APR-1996; 96US-0630405.
PR   07-JUN-1995; 95US-0487032.
XX
PA   (ASTR) ASTRA AB.
XX
PI   Berglindh OT, Smith D, Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20335.
XX
PS   Claim 9; Page -; 1481pp; English.
XX
CC   Helicobacter pylori nucleic acid sequences and related
CC   polypeptide(s) - useful for vaccines to treat or prevent H. pylori
CC   infection, and to detect Helicobacter
XX
SQ   Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;

Qy   572 aaagcccgcttcaaccctt 589
         |||||||||||||||||||
Db   311 AAAGCCGCTTCACCCTTT 294

RESULT 18
ID   AAT77453/c
AC   AAT77453;
XX
DT   11-AUG-1997 (first entry)
XX
DE   H. Pylori cytoplasmic protein ORF 24824087.aa.
XX
KW   Chronic gastritis; duodenal ulcer disease; activator;
KW   inhibitor; bacterial life cycle; vaccine; immunisation; detection;
KW   antisense; inhibition; cytoplasmic; Na+/H+ antiporter;
XX
OS   Escherichia coli; ds.
XX
OS   Helicobacter pylori.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..531

```
FT                /*tag= a
FT                /transl_except= (pos: 460..462, aa: Xaa)
FT                /transl_except= (pos: 520..522, aa: Xaa)
FT                /transl_except= (pos: 526..528, aa: Xaa)
FT                /note= "Xaa = Unknown"
XX
PN   WO9719098-A1.
XX
PN
PD   29-MAY-1997.
XX
PF   15-NOV-1996; 96WO-US18542.
XX
PR   17-NOV-1995; 95US-0561469.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Smith DH;
XX
DR   WPI: 1997-298052/27.
DR   P-PSDB; AAW24635.
XX
PT   Helicobacter pylori nucleic acid sequences and related proteins -
PT   used for diagnostics and therapeutics
XX
PS   Claim 1; Page 100; 235pp; English.
XX
CC   The present sequence encodes a Helicobacter pylori cytoplasmic
CC   protein, which was found to be homologous to Escherichia coli
CC   Na+/H+ antiporter protein following BLAST protein analysis.
CC   H. pylori has been strongly linked to chronic gastritis and
CC   duodenal ulcer disease. The nucleic acid sequences of the invention
CC   are used to evaluate compounds, especially activators or inhibitors
CC   of bacterial life cycle, for the ability to bind an H. pylori
CC   nucleic acid sequence. The nucleic acid sequences, and
CC   corresponding proteins, are also useful for generating vaccines for
CC   immunising subjects against H. pylori or for use in detecting the
CC   presence of Helicobacter species in a sample. Antisense nucleic
CC   acid sequences of these sequences are used to inhibit expression of
CC   a gene from Helicobacter species. H. pylori whole genomic DNA was
CC   isolated and nebulised to a median size of 2000 bp. Purified DNA
CC   fragments were blunt-ended and ligated to unique BstXI-linker
CC   adapters in 100-1000 fold molar excess. These linkers are
CC   complementary to the BstXI-cut pMPX vectors, while the overhang is
CC   not self-complementary. Therefore the linkers will not
CC   concatemerise nor will the cut vector re-ligate itself easily. The
CC   linker-adapter inserts were ligated to each of the 20 pMPX vectors
CC   to construct a series of shotgun subclone libraries. The purified
CC   DNA samples were then sequenced.
CC   Note: The ORF/protein reference number for this sequence was
CC   obtained from the related specification, WO9640893.
XX
SQ   Sequence 531 BP; 175 A; 88 C; 114 G; 151 T; 3 other;
```

```
Query Match             1.8%;  Score 18;  DB 18;  Length 531;
Best Local Similarity   100.0%;  Pred. No. 30;
  Matches  18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        572 aaagccgcttcacccttt 589
              ||||||||||||||||||
Db        311 AAAGCCGCTTCACCCTTT 294

RESULT 19
AAC44859
ID   AAC44859 standard; DNA; 791 BP.
XX
AC   AAC44859;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 44401.
XX
```

```
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128334.
PR   06-APR-1999;  99US-0128714.
PR   08-APR-1999;  99US-0128845.
PR   16-APR-1999;  99US-0130077.
PR   19-APR-1999;  99US-0130449.
PR   21-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136782.
PR   28-MAY-1999;  99US-0137222.
PR   01-JUN-1999;  99US-0137228.
PR   03-JUN-1999;  99US-0137502.
PR   04-JUN-1999;  99US-0137724.
PR   07-JUN-1999;  99US-0138094.
PR   08-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   10-JUN-1999;  99US-0139119.
PR   14-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   18-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
```

PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0145085.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145145.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0148958.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.

PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   28-SEP-1999;   99US-0156596.
PR   29-SEP-1999;   99US-0157117.
PR   04-OCT-1999;   99US-0157753.
PR   05-OCT-1999;   99US-0157865.
PR   06-OCT-1999;   99US-0158029.
PR   07-OCT-1999;   99US-0158232.
PR   08-OCT-1999;   99US-0158369.
PR   12-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   18-OCT-1999;   99US-0159638.
PR   21-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160801.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161361.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   26-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   28-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   29-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

```
Query Match         1.8%;   Score 18;  DB 21;  Length 791;
Best Local Similarity   100.0%;  Pred. No. 30;
Matches     18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    929 ggaagatggtttggaaga 946
          ||||||||||||||||||
Db    508 ggaagatggtttggaaga 525

RESULT   20
AAT67992/c
ID   AAT67992 standard; DNA; 1593 BP.
XX
AC   AAT67992;
XX
DT   16-JUL-1997   (first entry)
XX
DE   H. pylori cytoplasmic protein ORF 06cp30603orf11.
XX
KW   cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis; ds.
XX
OS   Helicobacter pylori.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             1..1593
FT                   /*tag= a
FT                   /note= "no stop codon given"
XX
PN   WO9640893-A1.
XX
PD   19-DEC-1996.
XX
PF   06-JUN-1996;   96WO-US09122.
XX
PR   01-APR-1996;   96US-0630405.
PR   07-JUN-1995;   95US-0487032.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Berglndh OT,  Smith D,  Mellgaerd BL;
XX
DR   WPI; 1997-052306/05.
DR   P-PSDB; AAW20739.
XX
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
XX
PS   Claim 9; Page 833; 141pp; English.
XX
CC   The present sequence encodes a H. pylori cytoplasmic protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
XX
SQ   Sequence 1593 BP; 541 A; 288 C; 334 G; 430 T; 0 other;

Query Match             1.8%;  Score 18;  DB 18;  Length 1593;
Best Local Similarity   100.0%;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
                        Pred. No. 30;

Qy       572 aaagccgcttcacccttt 589
             ||||||||||||||||||
Db      1376 AAAGCCGCTTCACCCTTT 1359

RESULT 21
ID   AAC77190 standard; cDNA; 1600 BP.
XX
AC   AAC77190;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF2745 polynucleotide sequence SEQ ID NO:5489.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000;   2000US-0540763.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42981.
XX
PA   (CURA-) CURAGEN CORP.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 4663-4664; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   anticonvulsant; antiarthritic; immunosuppressant;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1600 BP; 376 A; 435 C; 474 G; 315 T; 0 other;

Query Match             1.8%;  Score 18;  DB 21;  Length 1600;
Best Local Similarity   100.0%;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
                        Pred. No. 30;

Qy       464 gtcctgggcaccctggct 481
             ||||||||||||||||||
Db      1360 gtcctgggcaccctggct 1377

RESULT 22
     AAX08683/C
ID   AAX08683 standard; cDNA; 1626 BP.
XX
AC   AAX08683;
XX
DT   27-SEP-1999 (first entry)
```

DE   Novel nucleotide sequence encoding new protein (Clone AC222_1).
XX
KW   Polynucleotide; protein; nutrition; cytokine; cell proliferation;
KW   cell differentiation; immunostimulation; immunosuppression;
KW   haematopoiesis regulation; tissue growth; activin; inhibin;
KW   chemotaxis; chemokinesis; haemostasis; thrombolysis; receptor;
KW   ligand; anti-inflammatory; tumour suppression; gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   161..1129
FT   /*tag= a
FT   /product= "Novel protein"
XX
PN   WO9920644-A1.
XX
PD   29-APR-1999.
XX
PF   16-OCT-1998;   98WO-US22034.
XX
PR   18-OCT-1997;   97US-0955557.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Agostino MJ, Bowman MR, Evans C, Jacobs K, Lavallie ER;
PI   McCoy JM, Merberg D, Racie LA, Spaulding V, Treacy M;
XX
PS   Claim 14; Page 101-102; 136pp; English.
XX
PT   New polynucleotides encoding secreted human proteins
XX
DR   WPI; 1999-288272/24.
DR   P-PSDB; AAW85718.
XX
CC   The new human secreted proteins are encoded by polynucleotides
CC   obtained from human placenta, adult testes, fetal kidney, fetal
CC   brain, adult brain, adult brain and adult blood cDNA libraries.
CC   The polynucleotides and proteins are predicted to have biological
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals. Suggested
CC   activities include nutritional activity, cytokine and cell
CC   proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity. The polynucleotides are also stated to be useful for gene
CC   therapy. The sequences are identified by a secretory leader
CC   sequence motif in the polynucleotide and it is thought that the
CC   encoded proteins have biological activity by virtue of their secreted
CC   nature. This clone was designated AC222_1. A probe for this clone is
CC   described in AAX08698.
XX
SQ   Sequence 1626 BP; 560 A; 327 C; 291 G; 443 T; 5 other;

Qy   153 ggagtggcagaggatg 170
         ||||||||||||||||||
Db   930 GGGAGTGGCAGAGGGATG 913

Query Match   1.8%; Score 18; DB 20; Length 1626;
Best Local Similarity   100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 23
AAV17351
ID   AAV17351 standard; DNA; 2186 BP.
XX
AC   AAV17351;

DT   04-JUN-1998 (first entry)
XX
DE   Coding sequence for Kex2 protease inhibitor.
XX
KW   Kex2 protease; inhibitor; pesticide; kexstatin; ds.
XX
OS   Streptomyces platensis.
XX
PN   EP814160-A2.
XX
PD   29-DEC-1997.
XX
PF   19-JUN-1997;   97EP-0110075.
XX
PR   03-MAR-1997;   97JP-0048101.
PR   19-JUN-1996;   96JP-0158677.
PR   26-AUG-1996;   96JP-0224104.
XX
PA   (BIOM-) BIOMOLECULAR ENG RES INST.
XX
PI   Kikuchi N, Oda K, Shibano Y;
XX
DR   WPI; 1998-044337/05.
XX
PT   Kex2 protease inhibitor protein - with potential pharmaceutical and
PT   pesticidal applications
XX
PS   Disclosure; Page 18-19; 29pp; English.
XX
CC   This sequence is the coding sequence for an example of the inhibitor of
CC   the invention (the encoded protein is not given in the specification).
CC   The inhibitor is a Kex2 proteinase family enzyme inhibitor with a
CC   molecular weight of 11,500. The inhibitor/protein (termed kexstatin) is
CC   expected to have pharmaceutical and pesticidal applications.
XX
SQ   Sequence 2186 BP; 297 A; 827 C; 780 G; 282 T; 0 other;

Qy   463 ggtcctggggcaccctggc 480
         ||||||||||||||||||||
Db 1518 ggtcctggggcaccctggc 1535

Query Match   1.8%; Score 18; DB 19; Length 2186;
Best Local Similarity   100.0%; Pred. No. 30;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 24
AAZ09473/c
ID   AAZ09473 standard; DNA; 4120 BP.
XX
AC   AAZ09473;
XX
DT   02-NOV-1999 (first entry)
XX
DE   Human RNA helicase p135 DNA #2.
XX
KW   DEAD protein; insect cell vector; DEAD-superfamily; RNA helicase;
KW   DNA helicase; cancer research; cell line; inflammation; apoptosis;
KW   drug; anticancer; antiviral; p135; ds.
XX
OS   Homo sapiens.
XX
PN   WO9941390-A2.
XX
PD   19-AUG-1999.
XX
PF   01-FEB-1999;   99WO-EP00829.
XX
PR   12-FEB-1998;   98DE-1005781.
XX
PA   (AVET) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.

XX
PI Boehnisch B, Gallert K, Huels C, Muellner S;
XX
DR WPI; 1999-527373/44.
XX
PT New insect cell vector containing the sequence encoding a
PT DEAD-superfamily protein, particularly a nucleic acid helicase, used
PT e.g. for identifying potential pharmaceuticals
XX
PS Disclosure; Page 38-40; 43pp; German.
XX
CC This invention describes the construction of a novel insect cell vector
CC (A) which contains a nucleic acid (I) that codes for a protein (II) of
CC the DEAD-superfamily. (A), and recombinant insect viruses derived from
CC them, are used to express recombinant (II), particularly RNA and DNA
CC helicases. (II) are potentially useful for: (a) production of cell lines
CC for research into cancer, inflammation and apoptosis, or for clarifying
CC the mechanism of action of drugs, and (b) to identify pharmaceutical
CC activity in known compounds, e.g. anticancer and antiviral activities.
CC (II), which are difficult to express in insect cells, are expressed
CC at high level in insect cells, e.g. 300-400 mg per 109 cells. This
CC sequence encodes a human RNA-helicase p135 protein which is used in the
CC description of the method of the invention.
XX
SQ Sequence 4120 BP; 1099 A; 1021 C; 1236 G; 764 T; 0 other;

Qy 149 aggaggagtggcagagg 166
        |||||||||||||||||
Db 809 AGGAGGAGTGGCAGAGG 792

Query Match   1.8%;   Score 18;   DB 20;   Length 4120;
Best Local Similarity 100.0%;   Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 25
AAA60239/c
ID AAA60239 standard; DNA; 17 BP.
XX
AC AAA60239;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 60.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO20027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 3; Page 56; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human

CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 17 BP; 5 A; 2 C; 7 G; 3 T; 0 other;

Query Match   1.7%;   Score 17;   DB 21;   Length 17;
Best Local Similarity 100.0%;   Pred. No. 99;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 677 tcctgccaaatgcctc 693
        |||||||||||||||||
Db 17 TCCTGCCAAATGCCTTC 1

RESULT 26
AAA60241/c
ID AAA60241 standard; DNA; 17 BP.
XX
AC AAA60241;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 62.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO20027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 3; Page 56; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mn.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer patients, and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 17 BP; 3 A; 5 C; 6 G; 3 T; 0 other;

Query Match   1.7%;   Score 17;   DB 21;   Length 17;
Best Local Similarity 100.0%;   Pred. No. 99;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 824 gcacacctctggctgga 840
        |||||||||||||||||
Db 17 GCACACCTCTGGCTGGA 1

RESULT 27
ID   AAA60242 standard; DNA; 17 BP.
XX
AC   AAA60242;
XX
DT   07-DEC-2000 (first entry)
XX
DE   Human HPC2 cDNA sequencing primer SEQ ID NO: 63.
XX
KW   Human; mouse; prostate cancer predisposing gene; HPC2;
KW   human chromosome 17p; gene therapy; peptide therapy; drug design;
KW   PCR primer; sequencing primer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR   WPI; 2000-376481/32.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Example 3; Page 56; 157pp; English.
XX
CC   The present sequence is a primer used in the isolation of the human
CC   and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC   version of the gene is found on chromosome 17p. Some alleles cause a
CC   predisposition to cancer, particularly prostate cancer. This gene and its
CC   protein can be used in peptide and gene therapy for cancer patients, as
CC   well as being useful as diagnostic tools (both for cancer sufferers and
CC   those with a predisposition to the disease) and in the production of
CC   cancer drugs.
XX
SQ   Sequence 17 BP; 1 A; 5 C; 7 G; 4 T; 0 other;

Query Match   1.7%; Score 17; DB 21; Length 17;
Best Local Similarity   100.0%; Pred. No. 99;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   870 tgcgaggctctggtccg 886
              |||||||||||||||||
Db   1 tgcgaggctctggtccg 17

RESULT 28
ID   AAA32079 standard; DNA; 256 BP.
XX
AC   AAA32079;
XX
DT   05-JUL-2000 (first entry)
XX
DE   Plant microsatellite marker #1040.
XX
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Pinus radiata.

XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999;   99WO-NZ00092.
XX
PR   25-JUN-1998;   98US-0105307.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR   WPI; 2000-116958/10.
XX
PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
XX
PS   Claim 1; Page 382; 392pp; English.
XX
CC   Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences AAA32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for
CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 256 BP; 89 A; 41 C; 35 G; 91 T; 0 other;

Query Match   1.7%; Score 17; DB 21; Length 256;
Best Local Similarity   100.0%; Pred. No. 97;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   833 tgggctggaaagtggtct 849
              ||||||||||||||||||
Db   256 TGGCTGGAAAGTGGTCT 240

RESULT 29
ID   AAC32768/c
AC   AAC32768/c   AAC32768 standard; DNA; 292 BP.
XX
AC   AAC32768;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 580.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 14-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139492.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 21-JUN-1999; 99US-0139750.
PR 22-JUN-1999; 99US-0139763.
PR 23-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 24-JUN-1999; 99US-0140353.
PR 28-JUN-1999; 99US-0140354.
PR 29-JUN-1999; 99US-0140695.
PR 30-JUN-1999; 99US-0140823.
PR 01-JUL-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 02-JUL-1999; 99US-0141842.
PR 06-JUL-1999; 99US-0142154.
PR 08-JUL-1999; 99US-0142055.
PR 09-JUL-1999; 99US-0142390.
PR 12-JUL-1999; 99US-0142803.
PR 13-JUL-1999; 99US-0142920.
PR 14-JUL-1999; 99US-0142977.
PR 15-JUL-1999; 99US-0143542.
PR 16-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147004.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0153363.
PR 07-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0151139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0155458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157665.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159931.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.

Qy   623 gctccagcagtaccaca 639
         |||||||||||||||||
Db   242 GCTCCAGCAGTACCACA 226

Query Match             1.7%;  Score 17;  DB 21;  Length 292;
Best Local Similarity   100.0%;  Pred. No. 96;
Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

RESULT   30
AAA32050/c
ID   AAA32050 standard; DNA; 300 BP.
XX
AC   AAA32050;
XX
DT   05-JUL-2000   (first entry)
XX
DE   Plant microsatellite marker #1011.
XX
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Pinus radiata.
XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999; 99WO-NZ00092.
XX
PR   25-JUN-1998; 98US-0105307.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR   WPI; 2000-116958/10.
XX
XX
PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
```

```
PS   Claim 1; Page 373; 392pp; English.
XX
CC   Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences AAA32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for
CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 300 BP; 99 A; 45 C; 43 G; 113 T; 0 other;

Qy   833 tggctggaaagtggtct 849
         |||||||||||||||||
Db   292 TGGCTGGAAAGTGGTCT 276

Query Match             1.7%;  Score 17;  DB 21;  Length 300;
Best Local Similarity   100.0%;  Pred. No. 96;
Matches    17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

RESULT   31
AA242643/c
ID   AA242643 standard; cDNA; 330 BP.
XX
AC   AA242643;
XX
DT   01-FEB-2000   (first entry)
XX
DE   Human 5' EST isolated from a cDNA library SEQ ID NO:402.
XX
KW   Human; 5'; EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999; 99WO-IB00712.
XX
PR   09-APR-1998; 98US-0057719.
PR   28-APR-1998; 98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   P-PSDB; AAY65029.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS   Claim 1; Page 382; 837pp; English.
XX
CC   AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AA42265 to
CC   AA243052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (URBs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
```

CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAz42249 to AAz42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 330 BP; 76 A; 102 C; 86 G; 65 T; 1 other;

Qy 30 acctcatccacccggac 46
      |||||||||||||||||
Db 48 ACCTCATCCACCCGGAC 32

Query Match   1.7%; Score 17; DB 21; Length 330;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 32
AAT24629
ID AAT24629 standard; cDNA to mRNA; 347 BP.
XX
AC AAT24629;
XX
DT 07-OCT-1996 (first entry)
XX
DE Human gene signature HUMGS06689.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1655; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 347 BP; 61 A; 102 C; 108 G; 75 T; 1 other;

Qy 147 ccaggaggagtggcag 163
       ||||||||||||||||
Db 15 ccaggaggagtggcag 31

Query Match   1.7%; Score 17; DB 16; Length 347;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 33
AAC07468
ID AAC07468 standard; cDNA; 350 BP.
XX
AC AAC07468;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 11543.
XX
OS Homo sapiens.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 11543; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 350 BP; 96 A; 82 C; 78 G; 94 T; 0 other;

Query Match   1.7%; Score 17; DB 21; Length 350;
Best Local Similarity 100.0%; Pred. No. 96;

Qy   158 tggcaggaggatgcat 174
          |||||||||||||||||
Db   91 tggcaggaggatgcat 107

Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 34
AAF64559
ID   AAF64559 standard; cDNA; 398 BP.
XX
AC   AAF64559;
XX
DT   09-APR-2001 (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 315.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO20010102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999; 99US-0142310.
PR   02-JUL-1999; 99US-0142311.
XX
PA   (CHIR) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI   Reinhard C, Randazzo F, Kennedy GC, Pot D, Lanson G, Drmanac R;
PI   Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI   Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 590; 1046pp; English.
XX
DR   WPI; 2001-091805/10.
XX

CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotide. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 398 BP; 65 A; 105 C; 153 G; 73 T; 2 other;

Query Match   1.7%; Score 17; DB 22; Length 398;
Best Local Similarity   100.0%;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   462 gggtcctgggcacctg 478
          |||||||||||||||||
Db   130 gggtcctgggcacctg 146

RESULT 35
AAA32069/c
ID   AAA32069 standard; DNA; 426 BP.
XX
AC   AAA32069;
XX
DT   05-JUL-2000 (first entry)
XX
DE   Plant microsatellite marker #1030.
XX
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Pinus radiata.
XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999; 99WO-NZ00092.
XX
PR   25-JUN-1998; 98US-0105307.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR   WPI; 2000-116958/10.
XX

PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
XX
PS   Claim 1; Page 379; 392pp; English.
XX

CC   Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences AAA32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for
CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 426 BP; 124 A; 81 C; 71 G; 150 T; 0 other;

Query Match   1.7%; Score 17; DB 21; Length 426;
Best Local Similarity   100.0%;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   833 tggctgaaagtcggtct 849
          |||||||||||||||||
Db   257 TGGCTGGAAAGTGGTCT 241

RESULT 36
AAA32052/c
ID   AAA32052 standard; DNA; 431 BP.
XX
AC   AAA32052;
XX
DT   05-JUL-2000 (first entry)
XX
DE   Plant microsatellite marker #1013.
XX
XX

```
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Pinus radiata.
XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999;   99WO-NZ00092.
XX
PR   16-APR-1999;   98US-0105307.
XX
PI   Havukkala IJ,  Bloksberg LN,  Glenn M;
XX
DR   WPI; 2000-116958/10.
XX
PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
XX
PS   Claim 1; Page 374; 392pp; English.
XX
CC   Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences AAA32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for
CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 431 BP; 129 A; 78 C; 71 G; 152 T; 1 other;

Query Match             1.7%;  Score 17;  DB 21;  Length 431;
Best Local Similarity   100.0%;  Pred. No. 96;
Matches     17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        833 tggctggaaagtggtct 849
              |||||||||||||||||
Db        262 TGGCTGGAAAGTGGTCT 246


RESULT 37
AC   AAC46096/c
ID   AAC46096 standard; DNA; 459 BP.
XX
AC   AAC46096;
XX
DT   18-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 48895.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
```

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0125548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 16-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

```
Qy       623 gctccagcagtaccaca 639
             ||||||||||||||||||
Db       179 GCTCCAGCAGTACCACA 163
```

```
Query Match                      1.7%;    Score   17;   DB 21;   Length 459;
Best Local Similarity      100.0%;    Pred. No. 96;
Matches      17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
RESULT   38
AAC45096/c
ID    AAC45096 standard; DNA; 523 BP.
XX
AC    AAC45096;
XX
DT    18-OCT-2000    (first entry)
XX
DE    Arabidopsis thaliana DNA fragment SEQ ID NO: 45285.
XX
KW    Hybridisation assay; genetic mapping; gene expression control;
KW    protein identification; signal transduction pathway;
KW    metabolic pathway; promoter; termination sequence; ss.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000; 2000EP-0301439.
XX
PR    25-FEB-1999; 99US-0121825.
PR    05-MAR-1999; 99US-0123180.
PR    09-MAR-1999; 99US-0123548.
PR    23-MAR-1999; 99US-0125788.
PR    25-MAR-1999; 99US-0126264.
PR    29-MAR-1999; 99US-0126785.
PR    01-APR-1999; 99US-0127462.
```

```
PR    04-OCT-1999;  99US-0157117.
PR    05-OCT-1999;  99US-0157753.
PR    06-OCT-1999;  99US-0157865.
PR    07-OCT-1999;  99US-0158029.
PR    08-OCT-1999;  99US-0158232.
PR    12-OCT-1999;  99US-0158369.
PR    13-OCT-1999;  99US-0159293.
PR    13-OCT-1999;  99US-0159294.
PR    13-OCT-1999;  99US-0159295.
PR    14-OCT-1999;  99US-0159329.
PR    14-OCT-1999;  99US-0159330.
PR    14-OCT-1999;  99US-0159331.
PR    14-OCT-1999;  99US-0159637.
PR    14-OCT-1999;  99US-0159638.
PR    18-OCT-1999;  99US-0159584.
PR    21-OCT-1999;  99US-0160741.
PR    21-OCT-1999;  99US-0160767.
PR    21-OCT-1999;  99US-0160768.
PR    21-OCT-1999;  99US-0160770.
PR    21-OCT-1999;  99US-0160814.
PR    21-OCT-1999;  99US-0160815.
PR    22-OCT-1999;  99US-0160980.
PR    22-OCT-1999;  99US-0160981.
PR    22-OCT-1999;  99US-0160989.
PR    25-OCT-1999;  99US-0161404.
PR    25-OCT-1999;  99US-0161405.
PR    26-OCT-1999;  99US-0161360.
PR    26-OCT-1999;  99US-0161361.
PR    28-OCT-1999;  99US-0161920.
PR    28-OCT-1999;  99US-0161922.
PR    28-OCT-1999;  99US-0161993.
PR    29-OCT-1999;  99US-0162142.
```

```
PR    13-JUL-1999;  99US-0145542.
PR    14-JUL-1999;  99US-0143624.
PR    15-JUL-1999;  99US-0144005.
PR    16-JUL-1999;  99US-0144085.
PR    16-JUL-1999;  99US-0144086.
PR    19-JUL-1999;  99US-0144325.
PR    19-JUL-1999;  99US-0144331.
PR    19-JUL-1999;  99US-0144332.
PR    19-JUL-1999;  99US-0144333.
PR    19-JUL-1999;  99US-0144334.
PR    19-JUL-1999;  99US-0144335.
PR    20-JUL-1999;  99US-0144352.
PR    20-JUL-1999;  99US-0144884.
PR    21-JUL-1999;  99US-0144814.
PR    21-JUL-1999;  99US-0145086.
PR    22-JUL-1999;  99US-0145088.
PR    22-JUL-1999;  99US-0145085.
PR    22-JUL-1999;  99US-0145087.
PR    22-JUL-1999;  99US-0145089.
PR    22-JUL-1999;  99US-0145192.
PR    23-JUL-1999;  99US-0145145.
PR    23-JUL-1999;  99US-0145218.
PR    23-JUL-1999;  99US-0145224.
PR    26-JUL-1999;  99US-0145276.
PR    27-JUL-1999;  99US-0145913.
PR    27-JUL-1999;  99US-0145919.
PR    28-JUL-1999;  99US-0145951.
PR    02-AUG-1999;  99US-0146386.
PR    02-AUG-1999;  99US-0146388.
PR    02-AUG-1999;  99US-0146389.
PR    03-AUG-1999;  99US-0147038.
PR    04-AUG-1999;  99US-0147204.
PR    04-AUG-1999;  99US-0147302.
PR    05-AUG-1999;  99US-0147192.
PR    05-AUG-1999;  99US-0147260.
PR    06-AUG-1999;  99US-0147303.
PR    06-AUG-1999;  99US-0147416.
PR    09-AUG-1999;  99US-0147493.
PR    09-AUG-1999;  99US-0147935.
PR    10-AUG-1999;  99US-0148171.
PR    11-AUG-1999;  99US-0148319.
PR    12-AUG-1999;  99US-0148341.
PR    13-AUG-1999;  99US-0148565.
PR    13-AUG-1999;  99US-0148684.
PR    16-AUG-1999;  99US-0149368.
PR    17-AUG-1999;  99US-0149175.
PR    18-AUG-1999;  99US-0149426.
PR    20-AUG-1999;  99US-0149722.
PR    20-AUG-1999;  99US-0149723.
PR    23-AUG-1999;  99US-0149929.
PR    23-AUG-1999;  99US-0149902.
PR    25-AUG-1999;  99US-0150566.
PR    26-AUG-1999;  99US-0150884.
PR    27-AUG-1999;  99US-0151065.
PR    27-AUG-1999;  99US-0151066.
PR    27-AUG-1999;  99US-0151080.
PR    30-AUG-1999;  99US-0151303.
PR    31-AUG-1999;  99US-0151438.
PR    01-SEP-1999;  99US-0151930.
PR    07-SEP-1999;  99US-0152363.
PR    10-SEP-1999;  99US-0153070.
PR    13-SEP-1999;  99US-0153758.
PR    15-SEP-1999;  99US-0154018.
PR    16-SEP-1999;  99US-0154039.
PR    20-SEP-1999;  99US-0154779.
PR    22-SEP-1999;  99US-0155139.
PR    23-SEP-1999;  99US-0155486.
PR    24-SEP-1999;  99US-0155659.
PR    28-SEP-1999;  99US-0156458.
PR    29-SEP-1999;  99US-0156596.
```

```
PR  06-APR-1999;  99US-0128234.
PR  08-APR-1999;  99US-0128714.
PR  16-APR-1999;  99US-0129845.
PR  19-APR-1999;  99US-0130077.
PR  21-APR-1999;  99US-0130449.
PR  23-APR-1999;  99US-0130510.
PR  23-APR-1999;  99US-0130891.
PR  28-APR-1999;  99US-0131449.
PR  30-APR-1999;  99US-0132048.
PR  30-APR-1999;  99US-0132407.
PR  04-MAY-1999;  99US-0132407.
PR  05-MAY-1999;  99US-0132484.
PR  06-MAY-1999;  99US-0132485.
PR  06-MAY-1999;  99US-0132486.
PR  07-MAY-1999;  99US-0132487.
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  14-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  14-JUN-1999;  99US-0138847.
PR  16-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  17-JUN-1999;  99US-0139453.
PR  18-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.

PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0145086.
PR  21-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147260.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147302.
PR  06-AUG-1999;  99US-0147303.
PR  09-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  10-AUG-1999;  99US-0147935.
PR  11-AUG-1999;  99US-0148171.
PR  12-AUG-1999;  99US-0148319.
PR  13-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  16-AUG-1999;  99US-0148684.
PR  17-AUG-1999;  99US-0149368.
PR  18-AUG-1999;  99US-0149175.
PR  20-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149929.
PR  25-AUG-1999;  99US-0149930.
PR  26-AUG-1999;  99US-0150566.
PR  27-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  30-AUG-1999;  99US-0151066.
PR  31-AUG-1999;  99US-0151080.
PR  01-SEP-1999;  99US-0151303.
PR  07-SEP-1999;  99US-0151438.
PR  10-SEP-1999;  99US-0151930.
PR  13-SEP-1999;  99US-0153070.
PR  15-SEP-1999;  99US-0153363.
PR  16-SEP-1999;  99US-0154018.
PR  20-SEP-1999;  99US-0154039.
PR  22-SEP-1999;  99US-0154779.
PR  23-SEP-1999;  99US-0155139.
PR  24-SEP-1999;  99US-0155486.
PR  28-SEP-1999;  99US-0155659.
PR  29-SEP-1999;  99US-0156458.
PR  04-OCT-1999;  99US-0156996.
PR  05-OCT-1999;  99US-0157117.
PR  06-OCT-1999;  99US-0157753.
PR  07-OCT-1999;  99US-0157865.
PR  08-OCT-1999;  99US-0158029.
PR  12-OCT-1999;  99US-0158232.
PR  13-OCT-1999;  99US-0158369.
                  99US-0159293.
```

```
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match    1.7%;   Score 17;   DB 21;   Length 523;
Best Local Similarity  100.0%;   Pred. No. 96;
Matches  17;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy  623 gctccagcagtaccaca 639
        |||||||||||||||||
Db  243 GCTCCAGCAGTACCACA 227


RESULT 39
ID   AAV29360/c standard; DNA; 559 BP.
AC   AAV29360;
XX
DT   31-JUL-1998  (first entry)
XX
DE   Calcium ion channel alpha1 subunit exons 33, 34/intron partial sequence.
XX
KW   Calcium ion channel alpha1 subunit; human; episodic ataxia type 2;
KW   familial hemiplegic migraine; FHM; EA-2; treatment; diagnosis;
KW   exon; intron; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   intron          1..156
FT                   /*tag= a
FT                   /number= 32
FT                   /note= "partial sequence"
FT   exon            157..222
FT                   /*tag= b
FT                   /number= 33
FT   intron          223..394
FT                   /*tag= c
FT                   /number= 33
FT   exon            395..509
FT                   /*tag= d
FT                   /number= 34
FT   intron          510..559
FT                   /*tag= e
FT                   /number= 34
FT                   /note= "partial sequence"
XX
PN   EP834561-A1.
```

```
XX
PD   08-APR-1998.
XX
PF   27-SEP-1996;   96EP-0202707.
XX
PR   27-SEP-1996;   96EP-0202707.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
XX
PI   Ferrari MD, Frants RRIE, Ophoff RA, Terwindt GM;
XX
DR   WPI; 1998-195461/18.
XX
PT   New human nucleic acid associated with migraine and episodic ataxia
PT   type 2 - useful for diagnosis and development of specific treatments
XX
PS   Disclosure; Fig 1; 157pp; English.
XX
CC   Sequences shown in AAV29330 to AAV29371 represent the 47 exons and
CC   flanking intronic sequences containing the complete coding region of the
CC   human calcium ion channel alpha 1 subunit gene and part of untranslated
CC   sequences. The channel is related to familial hemiplegic migraine (FHM)
CC   and/or episodic ataxia type 2 (EA-2) and is derived from, related to or
CC   associated with a gene present in humans on chromosome 19p13.1-13.2. The
CC   encoding gene can be used to localise or identify genes related to
CC   episodic neurological disorders, specifically migraine, FHM or EA-2, but
CC   also epilepsy. The isolated or a recombinant nucleic acid can also be
CC   used to distinguish between alleles of the corresponding gene. Cells and
CC   animals containing recombinant expression vectors comprising the nucleic
CC   acid can be useful in study, development and treatment of migraine, FHM,
CC   EA-2 and epilepsy. Proteins or peptides encoded by the nucleic acid and
CC   natural or synthetic antibodies against the proteins can be used to
CC   diagnose FHM, EA-2, migraine and other neurological conditions
CC   associated with cation channel disfunction.
XX
SQ   Sequence 559 BP; 119 A; 157 C; 160 G; 120 T; 3 other;

Query Match    1.7%;   Score 17;   DB 19;   Length 559;
Best Local Similarity  100.0%;   Pred. No. 96;
Matches  17;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy   18 aaacccagctcaacctc 34
         |||||||||||||||||
Db  339 AAACCCAGCTCAACCTC 323


RESULT 40
ID   AAF64735 standard; cDNA; 694 BP.
AC   AAF64735;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 491.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000;   2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
```

XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac R, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 614-615; 1046pp; English.
XX

CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 694 BP; 127 A; 176 C; 246 G; 140 T; 5 other;

Qy   462 gggtcctggcaccctg 478
              |||||||||||||||||
Db   134 gggtcctggcaccctg 150

Query Match   1.7%; Score 17; DB 22; Length 694;
Best Local Similarity 100.0%; Pred. No. 96;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 41
ID   AAX24979/c
     AAX24979 standard; cDNA; 791 BP.
XX
AC   AAX24979;
XX
DT   05-JUL-1999 (first entry)
XX
DE   Clone GJ156 encoding TRAIN-R secreted form C-terminus.
XX
KW   TRAIN-R; receptor; human; tumour necrosis factor receptor;
KW   agonist; antagonist; cancer; immunological disease; therapy;
KW   cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Intron   1..350
FT   /tag= a
FT   exon   351..790
FT   /tag= b
FT   CDS   352..444
FT   /tag= c
FT   /partial
FT   3'UTR   45..790
FT   /product= "TRAIN-R secreted form C-terminus"
FT   /tag= d
XX
PN   WO9913078-A1.
XX
PD   18-MAR-1999.

XX
PF   11-SEP-1998; 98WO-US19030.
XX
PR   06-MAY-1998; 98US-0084422.
PR   12-SEP-1997; 97US-0058631.
XX
PA   (BIOJ) BIOGEN INC.
XX
PI   Hession C, Tschopp J;
XX
DR   WPI; 1999-229238/19.
DR   P-PSDB; AAW98147.
XX
PT   New cysteine-rich tumor necrosis factor receptor
XX
PS   Claim 1; Page 28; 30pp; English.
XX

CC The present sequence includes an exon encoding the C-terminus (see
CC AAW98147) of a soluble form of a novel human cysteine-rich tumour
CC necrosis factor receptor family member termed TRAIN-R. It comprises
CC clone GJ156, obtained from a Clontech human adult lung cDNA library.
CC The encoded 30-amino acid C-terminal peptide is identical to amino
CC acids 121-149 of the composite TRAIN-R protein given in AAW98146 and
CC to amino acids 121-150 of the C-terminus of murine TRAIN-R short
CC form (secreted protein, see AAW98144). The soluble protein is
CC expected to inhibit signalling by the full-length TRAIN-R. Human
CC TRAIN-R is expressed at low levels in every tissue and cell line
CC tested thus far, with higher expression detected in heart, prostate,
CC ovary, testis, peripheral blood lymphocytes, thyroid and adrenal
CC gland. Cell death can be induced by administering an agent capable
CC of inhibiting the binding of TRAIN-R to its ligand. A claimed method
CC of treating, or reducing, the advancement, severity or effects of
CC an immunological disease in a mammal comprises administering a
CC pharmaceutical composition which comprises a TRAIN-R blocking agent,
CC e.g. soluble TRAIN-R. TRAIN-R can be fused to an immunoglobulin to
CC produce a fusion protein which may be targeted to various sites.
CC It can be used in binding assays, and to identify antagonists and
CC agonists. Anti-TRAIN-R antibodies can be used to reduce the
CC severity of an immune response or to treat cancer. TRAIN-R
CC blocking agents can also be used to reduce the severity or effects
CC of an immunological disease (all claimed).
XX
SQ Sequence 791 BP; 202 A; 189 C; 165 G; 235 T; 0 other;

Qy   942 gaagaggaagcagtgga 958
              |||||||||||||||||
Db   203 GAAGAGGAAGCAGTGGA 187

Query Match   1.7%; Score 17; DB 20; Length 791;
Best Local Similarity 100.0%; Pred. No. 95;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 42
ID   AAZ91823
     AAZ91823 standard; DNA; 855 BP.
XX
AC   AAZ91823;
XX
DT   02-JUN-2000 (first entry)
XX
DE   Streptococcus pneumoniae DNA sequence ID33.
XX
KW   Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW   bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW   kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW   pneumococcal septicaemia; sinusitis; meningitis; therapy; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200006738-A2.
XX

PD　10-FEB-2000.
XX
PF　27-JUL-1999.;　　99WO-GB02452.
XX
PR　27-JUL-1998;　98GB-0016336.
PR　19-MAR-1999;　99US-0125329.
XX
PA　(MICR-) MICROBIAL TECHNICS LTD.
XX
PI　Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX
XX
PI
DR　WPI; 2000-195301/17.
DR　P-PSDB; AAX81727.
XX
PS　Claim 2; Page 47-48; 76pp; English.
XX
PT　Streptococcal proteins and polynucleotides useful for diagnosis,
PT　treatment and prophylaxis of bacterial infections -
XX
CC　This sequence encodes a Streptococcus pneumoniae protein of the
CC　invention. The proteins (or their homologues, derivatives and/or
CC　fragments)are useful as immunogens or antigens. Immunogenic or antigenic
CC　compositions comprising the proteins are useful as vaccines and also in
CC　diagnostic assays. The sequences are useful for the detection or
CC　diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC　with them. Agents capable of antagonising, inhibiting or interfering with
CC　the function or expression of the protein or polypeptide are useful in
CC　medical compositions in the treatment or prophylaxis of S. pneumoniae
CC　infection. As the sequences can be used to treat bacterial pneumonia,
CC　they can be used to treat bacterial pneumonia, which has high rates in
CC　young children, the elderly, and in patients with predisposing conditions
CC　such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC　or with immunosuppressive disorders, especially AIDS. They can also be
CC　used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC　meningitis.
XX
SQ　Sequence 855 BP; 235 A; 173 C; 202 G; 245 T; 0 other;

Query Match　　1.7%;　Score 17;　DB 21;　Length 855;
Best Local Similarity　100.0%;
Matches　17;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　759 gatttggaagagtttca 775
　　　|||||||||||||||||
Db　139 gatttggaagagtttca 155

RESULT　43
ID　AAX40199 standard; DNA; 1022 BP.
XX
AC　AAX40199;
XX
DT　02-JUL-1999　(first entry)
XX
DE　MAGE-4 encoding gene.
XX
KW　Cancer associated antigen; diagnosis; research; treatment; human;
KW　breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW　prostate cancer; ss.
XX
OS　Homo sapiens.
XX
PN　WO9904265-A2.
XX
PD　28-JAN-1999.
XX
PF　15-JUL-1998;　98WO-US14679.
XX
PR　22-JUN-1998;　98US-0102322.
PR　17-JUL-1997;　97US-0896164.
PR　10-OCT-1997;　97US-0061599.

PR　10-OCT-1997;　97US-0061765.
PR　10-OCT-1997;　97US-0948705.
PR　11-OCT-1997;　97GB-0021697.
XX
PA　(LUDW-) LUDWIG INST CANCER RES.
XX
PI　Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI　Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI　Tureci O;
XX
XX
DR　WPI; 1999-132448/11.
DR　P-PSDB; AAY06998.
XX
PS　Claim 67; Page 780; 787pp; English.
XX
PT　New isolated cancer associated nucleic acids and polypeptides -
PT　isolated using sera from cancer patients, used to develop products
PT　for the diagnosis, monitoring or treatment of cancers
XX
CC　The invention relates to a method for diagnosing a disorder characterised
CC　by expression of a human cancer associated antigen precursor coded for by
CC　a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC　biological sample isolated from a subject with an agent that specifically
CC　binds to the NAM, an expression product or a fragment of an expression
CC　product complexed with an HLA molecule; and (b) determining the
CC　interaction between the agent and the NAM or the expression product as a
CC　determination of the disorder. The products and methods can be used in
CC　the diagnosis, monitoring, research, or treatment of conditions
CC　characterised by the expression of various cancer associated antigens.
CC　The invention provides nucleic acid sequences and encoded polypeptides
CC　which are cancer associated antigen precursors expressed in human breast
CC　cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC　lung cancer.
XX
SQ　Sequence 1022 BP; 230 A; 273 C; 302 G; 217 T; 0 other;

Query Match　　1.7%;　Score 17;　DB 20;　Length 1022;
Best Local Similarity　100.0%;
Matches　17;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　463 ggtcctgggcaccctgg 479
　　　|||||||||||||||||
Db　203 ggtcctgggcaccctgg 219

RESULT　44
ID　AAQ67866 standard; DNA; 1084 BP.
XX
AC　AAQ67866;
XX
DT　22-MAR-1995　(first entry)
XX
DE　H6/MAGE-1 expression cassette from pMAW037.
XX
KW　Polymerase chain reaction; primer; amplify; NYVAC; ALVAC; recombinant;
KW　human; MAGE-1; melanoma-associated antigen; M22-E; testis; pTZ18RMAGE1;
KW　primary melanoma tumour cell; melanoma-derived cell line; tumour;
KW　poxvirus; antigenic response; immunological response; pathogen; ss.
XX
OS　Synthetic.
XX
FH　Key　　　　　　　　Location/Qualifiers
FH
FT　misc_feature　　　1..51
FT　　　　　　　　　　/*tag= a
FT　　　　　　　　　　/note= "Flanking sequence"
FT　promoter　　　　　52..178
FT　　　　　　　　　　/*tag= b
FT　　　　　　　　　　/note= "Vaccinia H6 promoter"
FT　CDS　　　　　　　　179..1009
FT　　　　　　　　　　/*tag= c
FT　　　　　　　　　　/product= MAGE-1

```
FT   misc_feature   1010..1084
FT                  /*tag= d
FT                  /note= "Flanking sequence"
XX
PN   WO9416716-A.
XX
PD   04-AUG-1994.
XX
PF   21-JAN-1994;   94WO-US00888.
XX
PR   21-JAN-1993;   93US-0007115.
PR   19-JAN-1994;   94US-0184009.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Cox WI, Paoletti E, Tartaglia J;
XX
DR   WPI; 1994-263767/32.
XX
PT   Attenuated recombinant virus used for cancer therapy - comprises
PT   DNA encoding cytokine and/or tumour associated antigen
XX
PS   Example 16; Fig 20; 232pp; English.
XX
CC   The sequences given in AAQ67865-66 represent expression cassettes
CC   containing the vaccinia H6 promoter and the human MAGE-1 gene which
CC   encodes human melanoma-associated antigen MZ2-E, in vCP235 and pMAW037,
CC   respectively. These sequences were used in the construction of NYVAC-
CC   and ALVAC-based recombinant viruses containing the MAGE-1 gene. MAGE-1
CC   is expressed in primary melanoma tumour cells, melanoma-derived cell
CC   lines and certain tumours of non-melanoma origins but not in normal
CC   cells except in testis. A first PCR fragment containing the last 18 bp
CC   and the initial 24 nucleotides of the MAGE-1 gene was generated and
CC   ligated to a second PCR fragment amplified from plasmid pTZ18RMAGE1
CC   which contains the initial 546 bp of the MAGE-1 coding sequence. The
CC   terminal sequence of MAGE-1 was amplified and a fusion product was
CC   generated containing the H6 promoter and the full length MAGE-1
CC   sequence. This construct may be introduced in to the poxvirus derived
CC   plasmids, ALVAC and NYVAC. The resulting viruses may be used in a
CC   composition for inducing an antigenic or immunological response, ie. for
CC   immunisation against pathogens.
XX
SQ   Sequence 1084 BP; 266 A; 256 C; 280 G; 282 T; 0 other;

Query Match    1.7%;   Score 17;   DB 15;   Length 1084;
Best Local Similarity   100.0%;   Pred. No. 95;
Matches  17;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy   463 ggtcctgggcaccctgg 479
         |||||||||||||||||
Db   292 ggtcctgggcaccctgg 308
```

```
RESULT 45
AZ08442
ID   AZ08442 standard; DNA; 1084 BP.
XX
AC   AZ08442;
XX
DT   19-OCT-1999   (first entry)
XX
DE   H6/MAGE-1 expression cassette and flanking regions from pMAW037.
XX
KW   Attenuated recombinant virus; cytokine; tumour associated antigen;
KW   NYVAC recombinant virus; ALVAC recombinant virus; gene therapy; rabies;
KW   cancer; tumour necrosis factor; nuclear phosphoprotein; p53; IL-2; GMCSF;
KW   interleukin; interferon; IFN-gamma; IL-4; melanoma associated antigen;
KW   carcinoembryonic antigen; immunisation; antigenic; poxvirus; influenza;
KW   immunological response; immunotherapy; vaccine; Newcastle Disease; ss.
XX
OS   Homo sapiens.
OS
OS   Synthetic.
```

```
OS   Vaccinia virus.
XX
PN   US5942235-A.
XX
PD   24-AUG-1999.
XX
PF   02-JUN-1995;   95US-0458356.
XX
PR   02-JUN-1995;   95US-0458356.
PR   24-DEC-1981;   81US-0334456.
PR   08-DEC-1982;   82US-0446824.
PR   19-JUN-1984;   84US-0622135.
PR   27-AUG-1984;   84US-0090209.
PR   28-AUG-1987;   87US-0090711.
PR   20-OCT-1987;   87US-0110335.
PR   25-APR-1988;   88US-0186054.
PR   23-AUG-1988;   88US-0234390.
PR   08-MAR-1989;   89US-0320471.
PR   14-FEB-1990;   90US-0478179.
PR   14-JUN-1990;   90US-0537890.
PR   07-JAN-1991;   91US-0638080.
PR   07-MAR-1991;   91US-0666056.
PR   11-JUN-1991;   91US-0713967.
PR   16-DEC-1991;   91US-0805567.
PR   03-MAR-1992;   92US-0847977.
PR   06-MAR-1992;   92US-0847951.
PR   04-MAY-1992;   92US-0881995.
PR   22-JUL-1992;   92US-0912278.
PR   20-JAN-1993;   93US-0007115.
PR   19-JAN-1994;   94US-0184009.
PR   14-APR-1994;   94US-0229926.
PR   13-SEP-1994;   94US-0306259.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Paoletti E;
XX
DR   WPI; 1999-493494/41.
XX
PT   Recombinant poxviruses comprising exogenous DNA encoding antigenic
PT   determinants useful in immunotherapy to immunize against cancers and
PT   other diseases such as influenza, Newcastle Disease and rabies
XX
PS   Example 16; Fig 20; 163pp; English.
XX
CC   The present invention describes a recombinant poxvirus (I), comprising
CC   exogenous DNA encoding an antigenic determinant of a pathogen which is
CC   then expressed in vivo in infected host cells after administration to a
CC   patient and therefore induces an immunological response. (I) may be used
CC   to vaccinate patients against a wide range of diseases and disorders
CC   depending on the type of antigen encoded by the exogenous DNA. (I) may
CC   be used to vaccinate against diseases such as rabies, influenza and
CC   Newcastle Disease. It is particularly useful for immunising against
CC   cancers. The poxvirus (I) also provides a means of manipulating
CC   lymphocytes and tumour cells for use in cell-based immunotherapeutic
CC   modalities for cancer. (I) also have enhanced safety compared to
CC   unattenuated viruses (attenuation reduces the virulence of the viruses)
CC   and known recombinant poxvirus vaccines. This increased level of safety
CC   reduces the possibility of a 'runaway' infection in the host and reduces
CC   the chance of transmission from vaccinated to unvaccinated individuals
CC   and contamination of the environment. The present sequence represents a
CC   H6/MAGE-1 expression cassette and flanking regions from pMAW037
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 1084 BP; 266 A; 256 C; 280 G; 282 T; 0 other;

Query Match    1.7%;   Score 17;   DB 20;   Length 1084;
Best Local Similarity   100.0%;   Pred. No. 95;
Matches  17;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy   463 ggtcctgggcaccctgg 479
```

Db   292 ggtcctgggcaccctgg 308
          ||||||||||||||||||

Search completed: October 28, 2001, 22:35:36
Job time: 2629 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:50:02

Title: SEQ1INS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacaagattcaaa.........aagccatcagcgtgggatg 1001

Scoring table: OLIGO_NUC
                 Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size :   8

Total number of hits satisfying chosen parameters: 255877

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

; Search time 51.72 Seconds
; (without alignments)
; 363.968 Million cell updates/sec

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	1.8	2186	2	US-08-878-546-9	Sequence 9, Appli
c	2	17	1.7	289	2	US-08-967-101-23	Sequence 23, Appl
c	3	17	1.7	289	2	US-08-592-541-23	Sequence 23, Appl
c	4	17	1.7	289	3	US-09-124-698-23	Sequence 23, Appl
c	5	17	1.7	289	4	US-09-127-480-23	Sequence 23, Appl
c	6	17	1.7	289	4	US-08-496-841C-23	Sequence 23, Appl
c	7	17	1.7	1084	2	US-08-184-009-110	Sequence 110, App
c	8	17	1.7	1084	2	US-08-458-356-110	Sequence 110, App
c	9	17	1.7	1084	2	US-08-184-009-109	Sequence 109, App
c	10	17	1.7	1094	2	US-08-458-356-109	Sequence 109, App
c	11	17	1.7	1330	4	US-09-118-442-29	Sequence 29, Appl
	12	17	1.7	1691	2	US-09-993-118-8	Sequence 8, Appl
	13	17	1.7	1691	1	US-08-845-528C-8	Sequence 8, Appli
	14	17	1.7	2419	1	US-07-807-043B-7	Sequence 7, Appli
	15	17	1.7	2419	1	US-08-299-849B-7	Sequence 7, Appl
	16	17	1.7	2419	3	US-08-142-368A-7	Sequence 7, Appl
	17	17	1.7	2419	3	US-08-967-727-7	Sequence 7, Appl
	18	17	1.7	2419	4	US-09-037-230D-7	Sequence 7, Appl
	19	17	1.7	2420	1	US-08-465-167A-23	Sequence 23, Appl
	20	17	1.7	4488	1	US-08-441-430-1	Sequence 1, Appl
	21	17	1.7	5674	1	US-07-807-043B-8	Sequence 8, Appli
	22	17	1.7	5674	1	US-08-190-411A-1	Sequence 1, Appl
	23	17	1.7	5674	1	US-08-299-849B-8	Sequence 8, Appl
	24	17	1.7	5674	2	US-08-560-024-1	Sequence 1, Appl
	25	17	1.7	5674	2	US-08-142-368A-8	Sequence 8, Appl
	26	17	1.7	5674	3	US-08-967-727-8	Sequence 8, Appl
	27	17	1.7	5674	4	US-08-037-230D-8	Sequence 8, Appli
c	28	17	1.7	11236	1	US-07-853-913-1	Sequence 1, Appli
c	29	17	1.7	68750	3	US-09-335-409-1	Sequence 1, Appl
c	30	16	1.7	76	1	US-07-753-110B-12	Sequence 12, Appl
c	31	16	1.6	76	1	US-08-503-730-16	Sequence 16, Appl
c	32	16	1.6	76	1	US-08-507-634-13	Sequence 13, Appl
c	33	16	1.6	235	4	US-08-505-223-207	Sequence 207, Appl
c	34	16	1.6	252	2	US-08-630-822A-97	Sequence 97, Appl
c	35	16	1.6	252	4	US-09-005-069-97	Sequence 97, Appl
c	36	16	1.6	252	3	US-08-906-769-104	Sequence 104, App
c	37	16	1.6	252	3	US-08-906-616-104	Sequence 104, App
c	38	16	1.6	252	3	US-08-817-795-104	Sequence 104, App
c	39	16	1.6	252	3	US-08-639-075A-104	Sequence 104, App
c	40	16	1.6	252	4	US-09-012-431-104	Sequence 104, App
c	41	16	1.6	252	4	US-09-012-692-104	Sequence 104, App
c	42	16	1.6	252	4	US-08-906-613-104	Sequence 104, App
c	43	16	1.6	252	5	PCT-US95-14442A-104	Sequence 104, App
c	44	16	1.6	291	4	US-09-060-756-254	Sequence 254, App
c	45	16	1.6	489	1	US-08-334-254-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-878-546-9
Sequence 9, Application US/08878546
Patent No. 5952463
GENERAL INFORMATION:
APPLICANT: SHIBANO, YUJI
APPLICANT: KIKUCHI, NORIHISA
APPLICANT: ODA, KOHEI
TITLE OF INVENTION: NOVEL PROTEINASE INHIBITOR AND
TITLE OF INVENTION: GENE ENCODING THE INHIBITOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: STEINBERG, RASKIN & DAVIDSON P.C.
STREET: 1140 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,546
FILING DATE: 19-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 158677/1996
FILING DATE: 19-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 224104/1996
FILING DATE: 26-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 48101/1997
FILING DATE: 03-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: DAVIDSON, CLIFFORD M.
REGISTRATION NUMBER: 32,728
REFERENCE/DOCKET NUMBER: 382.1009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)-768-3800
TELEFAX: (212)382-2124
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2186 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
;   ORGANISM: STREPTOMYCES PLATENSIS
;   STRAIN: Q268
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1477..1911
US-08-878-546-9

Qy   463 ggtcctgggcaccctggc 480
            ||||||||||||||||||
Db   1518 GGTCCTGGGCACCCTGGC 1535

RESULT 2
US-08-967-101-23/c
Sequence 23, Application US/08967101
Patent No. 5840540
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,101
FILING DATE: 10-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,541
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-967-101-23

Query Match   1.7%;   Score 17;   DB 2;   Length 289;
Best Local Similarity 100.0%; Pred. No. 34;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   943 aagaggaagcagtggaa 959
            |||||||||||||||||
Db   267 AAGAGGAAGCAGTGGAA 251

RESULT 3
US-08-592-541-23/c
Sequence 23, Application US/08592541
Patent No. 5986054
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,541
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-592-541-23

Query Match   1.7%;   Score 17;   DB 2;   Length 289;
Best Local Similarity 100.0%; Pred. No. 34;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   943 aagaggaagcagtggaa 959
            |||||||||||||||||
Db   267 AAGAGGAAGCAGTGGAA 251

RESULT 4
US-09-124-698-23/c
Sequence 23, Application US/09124698
Patent No. 6117978
GENERAL INFORMATION:
APPLICANT: ST. GEORGE-HYSLOP, PETER H
APPLICANT: ROMMENS, JOHANNA M
APPLICANT: FRASER, PAUL E
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 183
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: High Street Tower - 125 High Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```
;
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-124-698-23

Qy  943  aagaggaagcagtggaa  959
         |||||||||||||||||
Db  267  AAGAGGAAGCAGTGGAA  251

Query Match        1.7%;   Score 17;   DB 3;   Length 289;
Best Local Similarity   100.0%;   Pred. No. 34;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 5
US-09-127-480-23/c
Sequence 23, Application US/09127480
Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: High Street Tower - 125 High Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/127,480
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-23

Qy  943  aagaggaagcagtggaa  959
         |||||||||||||||||
Db  267  AAGAGGAAGCAGTGGAA  251

Query Match        1.7%;   Score 17;   DB 4;   Length 289;
Best Local Similarity   100.0%;   Pred. No. 34;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 6
US-08-496-841C-23/c
Sequence 23, Application US/08496841C
Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
;            ROMMENS, JOHANNA M
;            FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;            TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/496,841C
; FILING DATE: 28-Jun-1995
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul F. Fehlner, Ph.D.
; REGISTRATION NUMBER: 35,135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-08-496-841C-23

Qy  943  aagaggaagcagtggaa  959
         |||||||||||||||||
Db  267  AAGAGGAAGCAGTGGAA  251

Query Match        1.7%;   Score 17;   DB 4;   Length 289;
Best Local Similarity   100.0%;   Pred. No. 34;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

RESULT 7
US-08-184-009-110
Sequence 110, Application US/08184009
Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
```

Left page

```
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Curtis, Morris & Safford
   STREET: 530 Fifth Avenue
   CITY: New York
   STATE: NY
   COUNTRY: USA
   ZIP: 10036
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/184,009
   FILING DATE: 19-JAN-1994
   CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
   NAME: Frommer, William S.
   REGISTRATION NUMBER: 25,506
   REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 840-3333
   TELEFAX: (212) 840-0712
   TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 110:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1084 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: cDNA
US-08-184-009-110
```

```
Query Match   Best Local Similarity 100.0%;   1.7%;   Score 17;   DB 2;   Length 1084;
Matches 17;  Conservative 0;  Mismatches 0;  Pred. No. 34;  Indels 0;  Gaps 0;

Qy   463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db   292 GGTCCTGGGCACCCTGG 308
```

```
RESULT 8
US-08-458-356-110
Sequence 110, Application US/08458356
Patent No. 5942235
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Curtis, Morris & Safford
   STREET: 530 Fifth Avenue
   CITY: New York
   STATE: NY
   COUNTRY: USA
   ZIP: 10036
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/458,356
   FILING DATE: 02-JUN-1995
   CLASSIFICATION: 424
```

Right page

```
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 08/184,009
   FILING DATE: 19-JAN-1994
ATTORNEY/AGENT INFORMATION:
   NAME: Frommer, William S.
   REGISTRATION NUMBER: 25,506
   REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 840-3333
   TELEFAX: (212) 840-0712
   TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 110:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1084 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: cDNA
US-08-458-356-110
```

```
Query Match   Best Local Similarity 100.0%;   1.7%;   Score 17;   DB 2;   Length 1084;
Matches 17;  Conservative 0;  Mismatches 0;  Pred. No. 34;  Indels 0;  Gaps 0;

Qy   463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db   292 GGTCCTGGGCACCCTGG 308
```

```
RESULT 9
US-08-184-009-109
Sequence 109, Application US/08184009
Patent No. 5833975
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Curtis, Morris & Safford
   STREET: 530 Fifth Avenue
   CITY: New York
   STATE: NY
   COUNTRY: USA
   ZIP: 10036
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/184,009
   FILING DATE: 19-JAN-1994
   CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
   NAME: Frommer, William S.
   REGISTRATION NUMBER: 25,506
   REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 840-3333
   TELEFAX: (212) 840-0712
   TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 109:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1084 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: cDNA
US-08-184-009-109
```

Qy   463 ggtcctgggcaccctgg 479
             |||||||||||||||||
Db   314 GGTCCTCGGGCACCCTGG 330

Query Match   1.7%; Score 17; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
US-08-458-356-109
Sequence 109, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Curtis, Morris & Safford
;   STREET: 530 Fifth Avenue
;   CITY: New York
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 10036
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/458,356
;   FILING DATE: 02-JUN-1995
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/184,009
;   FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Frommer, William S.
;   REGISTRATION NUMBER: 25,506
;   REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 840-3333
;   TELEFAX: (212) 840-0712
;   TELEX: 42506GCURTMS
; SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1094 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-458-356-109

Qy   463 ggtcctgggcaccctgg 479
             ||||||||||||||||||
Db   314 GGTCCTCGGGCACCCTGG 330

Query Match   1.7%; Score 17; DB 2; Length 1094;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
US-09-118-442-29/C
Sequence 29, Application US/09118442B
; Patent No. 6197561
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu

; APPLICANT: Beach, Larry R.
; APPLICANT: Wang, Xun
; APPLICANT: Bowen, Benjamin A.
; TITLE OF INVENTION: Genes Controlling Phytate Metabolism in
; TITLE OF INVENTION: Plants and Uses Thereof
; FILE REFERENCE: 0706
; CURRENT APPLICATION NUMBER: US/09/118,442B
; CURRENT APPLICATION FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER APPLICATION FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER APPLICATION FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER APPLICATION FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Zea mays
US-09-118-442-29

Qy   334 catcccgatgaagattc 350
             |||||||||||||||||
Db   493 CATCCCGATGAAGATTC 477

Query Match   1.7%; Score 17; DB 4; Length 1330;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
US-08-993-118-8
Sequence 8, Application US/08993118
; Patent No. 5997872
; GENERAL INFORMATION:
; APPLICANT: LUCAS, Sophie;
; APPLICANT: DE SMET, Charles;
; APPLICANT: BOON-FALLEUR, Thierry
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10022
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/993,118
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/845,528
;   FILING DATE: April 25, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mary Anne Schofield
;   REGISTRATION NUMBER: 36,669
;   REFERENCE/DOCKET NUMBER: LUD 5455
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 688-9200
;   TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1691 base pairs

```
;   TYPE: nucleotides
;   STRANDEDNESS: single stranded
;   TOPOLOGY: linear
US-08-993-118-8

Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db    317 GGTCCTGGGCACCCTGG 333
```

RESULT 13
US-08-845-528C-8
; Sequence 8, Application US/08845528C
; Patent No. 6027924
; GENERAL INFORMATION:
;   APPLICANT: LUCAS, Sophie;
;   APPLICANT: DE SMET, Charles;
;   APPLICANT: BOON-FALLEUR, Thierry
;   TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;   TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;   TITLE OF INVENTION: THEREOF
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   COUNTRY: USA
;   ZIP: 10022
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/845,528C
;   FILING DATE: April 25, 1997
;   CLASSIFICATION: 4335
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Mary Anne Schofield
;   REGISTRATION NUMBER: 36,669
;   REFERENCE/DOCKET NUMBER: LUD 5455
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 688-9200
;   TELEFAX: (212) 838-3884
;   INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1691 base pairs
;   TYPE: nucleotides
;   STRANDEDNESS: single stranded
;   TOPOLOGY: linear
US-08-845-528C-8

Query Match   1.7%;   Score 17;   DB 2;   Length 1691;
Best Local Similarity 100.0%; Pred. No. 34;
Matches   17;   Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db   317 GGTCCTGGGCACCCTGG 333

RESULT 14
US-07-807-043B-7
; Sequence 7, Application US/0807043B
; Patent No. 5342774
; GENERAL INFORMATION:
```

```
;   APPLICANT: Boon, Thierry, Van den Eynde, Beno t
;   TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
;   TITLE OF INVENTION: Rejection Antigens and Uses Thereof
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Felfe & Lynch
;     STREET: 805 Third Avenue
;     CITY: New York City
;     STATE: New York
;     ZIP: 10022
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;     COMPUTER: IBM
;     OPERATING SYSTEM: PC-DOS
;     SOFTWARE: Wordperfect
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/807,043B
;     FILING DATE: 19911212
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/764,364
;     FILING DATE: 23-SEPTEMBER-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/728,838
;     FILING DATE: 9-JULY-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 07/705,702
;     FILING DATE: 23-May-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Hanson, No. 534274man D.
;     REGISTRATION NUMBER: 30,946
;     REFERENCE/DOCKET NUMBER: LUD 253.3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 688-9200
;     TELEFAX: (212) 838-3884
;   INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2419 base pairs
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: singular
;     TOPOLOGY: linear
;     MOLECULE TYPE: genomic DNA
US-07-807-043B-7

Query Match          1.7%;    Score 17;   DB 1;    Length 2419;
Best Local Similarity  100.0%;  Pred. No. 34;
Matches   17;   Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db    739 GGTCCTGGGCACCCTGG 755
```

```
Query Match          1.7%;    Score 17;   DB 3;    Length 1691;
Best Local Similarity  100.0%;  Pred. No. 34;
Matches   17;   Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db    317 GGTCCTGGGCACCCTGG 333
```

RESULT 14
US-07-807-043B-7
; Sequence 7, Application US/0780704B
; Patent No. 5342774
; GENERAL INFORMATION:

RESULT 15
US-08-299-849B-7
; Sequence 7, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
;   APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
;   APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;   APPLICANT: Chomez, Patrick
;   TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
;   TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
;   NUMBER OF SEQUENCES: 48
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   ZIP: 10022
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

RESULT 16
US-08-142-368A-7
Sequence 7, Application US/08142368A
Patent No. 5925729
GENERAL INFORMATION:
; APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
; APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
; APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 561201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA

Query Match
Best Local Similarity 100.0%; Score 17; DB 1; Length 2419;
Matches 17; Conservative 0; Pred. No. 34; Mismatches 0; Indels 0; Gaps 0;

Qy   463 ggtcctgggcaccctgg 479
         |||||||||||||||||
Db   739 GGTCCTGGGCACCCTGG 755

US-08-299-849B-7

RESULT 17
US-08-967-727-7
Sequence 7, Application US/08967727
Patent No. 6025474
GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,727
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/142,368A
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 592572 9man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5253.4-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA

US-08-142-368A-7

Query Match
Best Local Similarity 100.0%; Score 17; DB 2; Length 2419;
Matches 17; Conservative 0; Pred. No. 34; Mismatches 0; Indels 0; Gaps 0;

Qy   463 ggtcctgggcaccctgg 479
         |||||||||||||||||
Db   739 GGTCCTGGGCACCCTGG 755

RESULT 18
US-08-037-230D-7
Sequence 7, Application US/08037230D
Patent No. 6235525
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/037,230D
FILING DATE: 26-MARCH-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364

APPLICATION NUMBER: PCT/US92/04354
PRIOR APPLICATION DATA:
FILING DATE: 22-MAY-1992
APPLICATION NUMBER: 07/807,043
PRIOR APPLICATION DATA:
FILING DATE: 12-DECEMBER-1991
APPLICATION NUMBER: 07/764,365
PRIOR APPLICATION DATA:
FILING DATE: 23-SEPTEMBER-1991
APPLICATION NUMBER: 07/728,838
PRIOR APPLICATION DATA:
FILING DATE: 9-JULY-1991
APPLICATION NUMBER: 07/705,702
PRIOR APPLICATION DATA:
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6025474man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-967-727-7

Qy 463 ggtcctgggcaccctgg 479
        |||||||||||||||||
Db 739 GGTCCTGGGCACCCTGG 755

Query Match
Best Local Similarity   1.7%;   Score 17;   DB 3;   Length 2419;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
Pred. No. 34;

RESULT 19
US-08-465-167A-23
Sequence 23, Application US/08465167A
Patent No. 5750395
GENERAL INFORMATION:
APPLICANT: Fikes, John D.
APPLICANT: Livingston, Brian D.
APPLICANT: Sette, Alessandro D.
APPLICANT: Sidney, John C.
TITLE OF INVENTION: DNA ENCODING MAGE-1 C-TERMINAL
TITLE OF INVENTION: IMMUNOGENIC PEPTIDES (as amended)
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 98111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,167A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/103,623
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 14137-60-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:

FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6235525man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5353
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2419 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-037-230D-7

Qy 463 ggtcctgggcaccctgg 479
        |||||||||||||||||
Db 739 GGTCCTGGGCACCCTGG 755

Query Match
Best Local Similarity   1.7%;   Score 17;   DB 4;   Length 2419;
Matches   17;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
Pred. No. 34;

```
;   LENGTH: 2420 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 626..1552
US-08-465-167A-23

Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db    739 GGTCCTGGGCACCCTGG 755

Query Match                 1.7%;    Score 17;   DB 1;    Length 2420;
Best Local Similarity   100.0%;    Pred. No. 34;
Matches  17;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 20
US-08-441-430-1
;   Sequence 1, Application US/08441430
;   Patent No. 5681942
;   GENERAL INFORMATION:
;     APPLICANT: Buchwald, Manuel
;     APPLICANT: Strathdee, Craig A.
;     APPLICANT: Mevrick, Rachel
;     APPLICANT: Mathew, Christopher George Porter
;     TITLE OF INVENTION: Fanconi Anemia Type C Gene
;     NUMBER OF SEQUENCES: 73
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Richard J. Polley, Esq.
;       ADDRESSEE: Klarquist, Sparkman, Campbell, Leigh &
;       ADDRESSEE: Whinston, LLP
;       STREET: 121 S.W. Salmon, Suite 1600
;       CITY: Portland
;       STATE: Oregon
;       COUNTRY: U.S.A.
;       ZIP: 97204
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Disk, 3+-inch
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: MS DOS
;       SOFTWARE: Wordperfect 5.1/ASCII Text File
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/441,430
;       FILING DATE: May 15, 1995
;       CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: U.S. 07/876,285
;       FILING DATE: April 29, 1992
;       APPLICATION NUMBER: U.S. 07/918,313
;       FILING DATE: July 21, 1992
;       APPLICATION NUMBER: U.S. 08/003,963
;       FILING DATE: January 15, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Richard J. Polley, Esq.
;       REGISTRATION NUMBER: 28,107
;       REFERENCE/DOCKET NUMBER: 3812-42824
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (503) 226-7391
;       TELEFAX: (503) 228-9446
;   INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 4488 base pairs
;       TYPE: Nucleic Acid
;       STRANDEDNESS: Double stranded
;       TOPOLOGY: Linear
;     MOLECULE TYPE: cDNA to mRNA
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
;     ORIGINAL SOURCE:
```

```
;       ORGANISM: Homo sapiens
;       IMMEDIATE SOURCE:
;         LIBRARY: Human cDNA
;       POSITION IN GENOME: (of corresponding genomic gene)
;         CHROMOSOME/SEGMENT: 9q
;         MAP POSITION: 22.3
;       UNITS:
US-08-441-430-1

Qy    917 tgaagccacctggaag 933
          ||||||||||||||||
Db   4189 TGAAGCCACCTGGAAG 4205

Query Match                 1.7%;    Score 17;   DB 1;    Length 4488;
Best Local Similarity   100.0%;    Pred. No. 34;
Matches  17;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

RESULT 21
US-07-807-043B-8
;   Sequence 8, Application US/07807043B
;   Patent No. 5342774
;   GENERAL INFORMATION:
;     APPLICANT: Boon, Thierry, Van den Eynde, Beno t
;     TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
;     TITLE OF INVENTION: Rejection Antigens and Uses Thereof
;     NUMBER OF SEQUENCES: 16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Felfe & Lynch
;       STREET: 805 Third Avenue
;       CITY: New York City
;       STATE: New York
;       ZIP: 10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;       COMPUTER: IBM
;       OPERATING SYSTEM: PC-DOS
;       SOFTWARE: Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/07/807,043B
;       FILING DATE: 19911212
;       CLASSIFICATION: 424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/764,364
;       FILING DATE: 23-SEPTEMBER-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/728,838
;       FILING DATE: 9-JULY-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/705,702
;       FILING DATE: 23-May-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Hanson, No. 534274man D.
;       REGISTRATION NUMBER: 30,946
;       REFERENCE/DOCKET NUMBER: LUD 253.3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (212) 688-9200
;       TELEFAX: (212) 838-3884
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 5674 base pairs
;       TYPE: NUCLEIC ACID
;       STRANDEDNESS: singular
;       TOPOLOGY: linear
;     MOLECULE TYPE: genomic DNA
;     FEATURE:
;       NAME/KEY: MAGE-1 gene
US-07-807-043B-8

Query Match                 1.7%;    Score 17;   DB 1;    Length 5674;
Best Local Similarity   100.0%;    Pred. No. 34;
```

```
Qy      463  ggtcctgggcaccctgg 479
             |||||||||||||||||
Db     3994  GGTCCTGGGCACCCTGG 4010
```

Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 22
US-08-190-411A-1
; Sequence 1, Application US/08190411A
; Patent No. 554104
; GENERAL INFORMATION:
;   APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
;   APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
;   APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry;
;   APPLICANT: Old, Lloyd J.
;   TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
;   TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-1,
;   TITLE OF INVENTION: AND MAGE-1 DERIVED IMMUNOGENIC PEPTIDES
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   ZIP: 10022
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;   COMPUTER: IBM
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/190,411A
;   FILING DATE: 01-FEBRUARY-1994
;   CLASSIFICATION: 436
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 037,230
;   FILING DATE: 26-MARCH-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US92/04354
;   FILING DATE: 22-MAY-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/807,043
;   FILING DATE: 12-DECEMBER-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/764,364
;   FILING DATE: 23-SEPTEMBER-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/728,838
;   FILING DATE: 9-JULY-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/705,702
;   FILING DATE: 23-MAY-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hanson, No. 554104man D.
;   REGISTRATION NUMBER: 30,946
;   REFERENCE/DOCKET NUMBER: LUD 5354
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 688-9200
;   TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5674 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: genomic DNA
;   FEATURE:
;   NAME/KEY: MAGE-1 gene
US-08-190-411A-1

Query Match   1.7%; Score 17; DB 1; Length 5674;

```
Qy      463  ggtcctgggcaccctgg 479
             |||||||||||||||||
Db     3994  GGTCCTGGGCACCCTGG 4010
```

Best Local Similarity   100.0%; Pred. No. 34;
Matches   17; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 23
US-08-299-849B-8
; Sequence 8, Application US/08298849B
; Patent No. 5612201
; GENERAL INFORMATION:
;   APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
;   APPLICANT: Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;   APPLICANT: Chomez, Patrick
;   TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
;   TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
;   NUMBER OF SEQUENCES: 48
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Felfe & Lynch
;   STREET: 805 Third Avenue
;   CITY: New York City
;   STATE: New York
;   ZIP: 10022
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
;   COMPUTER: IBM
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/299,849B
;   FILING DATE: 1-SEPTEMBER-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/037,230
;   FILING DATE: 26-MARCH-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US92/04354
;   FILING DATE: 22-MAY-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/807,043
;   FILING DATE: 12-DECEMBER-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/764,364
;   FILING DATE: 23-SEPTEMBER-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/728,838
;   FILING DATE: 9-JULY-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/705,702
;   FILING DATE: 23-May-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Hanson, No. 5612201man D.
;   REGISTRATION NUMBER: 30,946
;   REFERENCE/DOCKET NUMBER: LUD 5355
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 688-9200
;   TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5674 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: genomic DNA
;   FEATURE:
;   NAME/KEY: MAGE-1 gene
US-08-299-849B-8

Query Match   1.7%; Score 17; DB 1; Length 5674;
Best Local Similarity 100.0%; Pred. No. 34;
```

Qy      463 ggtcctgggcaccctgg 479
            |||||||||||||||||
Db     3994 GGTCCTGGGCACCCTGG 4010

Matches    17;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

RESULT 24
US-08-560-024-1
Sequence 1, Application US/08560024
Patent No. 5843448
GENERAL INFORMATION:
  APPLICANT: Chen, Yao-Tseng; Stockert, Elisabeth;
  APPLICANT: Chen, Yachi; Garin-Chesa, Pilar; Rettig, Wolfgang J.;
  APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry;
  APPLICANT: Old, Lloyd J.
  TITLE OF INVENTION: MONOCLONAL ANTIBODIES WHICH BIND TO
  TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR MAGE-1, RECOMBINANT MAGE-1,
  TITLE OF INVENTION: AND MAGE-1 DERIVED IMMUNOGENIC PEPTIDES
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Felfe & Lynch
    STREET: 805 Third Avenue
    CITY: New York City
    STATE: New York
    ZIP: 10022
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
    COMPUTER: IBM
    OPERATING SYSTEM: PC-DOS
    SOFTWARE: Wordperfect
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/560,024
    FILING DATE:
  PRIOR APPLICATION DATA:
    CLASSIFICATION: 514
    APPLICATION NUMBER: US/08/190,411
    FILING DATE: 01-FEBRUARY-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 037,230
    FILING DATE: 26-MARCH-1993
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/US92/04354
    FILING DATE: 22-MAY-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/807,043
    FILING DATE: 12-DECEMBER-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/764,364
    FILING DATE: 23-SEPTEMBER-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/728,838
    FILING DATE: 9-JULY-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/705,702
    FILING DATE: 23-MAY-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Hanson, No. 5843448man D.
    REGISTRATION NUMBER: 30,946
    REFERENCE/DOCKET NUMBER: LUD 5354
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212) 688-9200
    TELEFAX: (212) 838-3884
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 5674 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: genomic DNA
  FEATURE:
    NAME/KEY: MAGE-1 gene

---

Query Match                       1.7%;    Score 17;    DB 2;    Length 5674;
Best Local Similarity    100.0%;    Pred. No. 34;
Matches    17;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy      463 ggtcctgggcaccctgg 479
            |||||||||||||||||
Db     3994 GGTCCTGGGCACCCTGG 4010

RESULT 25
US-08-142-368A-8
Sequence 8, Application US/08142368A
Patent No. 5925729
GENERAL INFORMATION:
  APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
  APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
  APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
  TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
  TITLE OF INVENTION: Rejection Antigens and Uses Thereof
  NUMBER OF SEQUENCES: 26
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Felfe & Lynch
    STREET: 805 Third Avenue
    CITY: New York City
    STATE: New York
    ZIP: 10022
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
    COMPUTER: IBM
    OPERATING SYSTEM: PC-DOS
    SOFTWARE: Wordperfect
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/142,368A
    FILING DATE: 02-MAY-1994
  PRIOR APPLICATION DATA:
    CLASSIFICATION: 435
    APPLICATION NUMBER: PCT/US92/04354
    FILING DATE: 22-MAY-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/807,043
    FILING DATE: 12-DECEMBER-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/764,364
    FILING DATE: 23-SEPTEMBER-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/728,838
    FILING DATE: 9-JULY-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/705,702
    FILING DATE: 23-MAY-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Hanson, No. 5925729man D.
    REGISTRATION NUMBER: 30,946
    REFERENCE/DOCKET NUMBER: LUD 5253.4-US
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212) 688-9200
    TELEFAX: (212) 838-3884
  INFORMATION FOR SEQ ID NO: 8:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 5674 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: genomic DNA
  FEATURE:
    NAME/KEY: MAGE-1 gene

```
Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db   3994 GGTCCTGGGCACCCTGG 4010
```

RESULT 26
US-08-967-727-8
; Sequence 8, Application US/08967727
; Patent No. 6025474
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,727
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,365
; FILING DATE: 23-SEPTEMBER-1991
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6025474man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene

US-08-967-727-8

Query Match   1.7%; Score 17; DB 3; Length 5674;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db   3994 GGTCCTGGGCACCCTGG 4010
```

RESULT 27
US-08-037-230D-8
; Sequence 8, Application US/08037230D
; Patent No. 6235525
; GENERAL INFORMATION:
; APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
; APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
; TITLE OF INVENTION: Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/037,230D
; FILING DATE: 26-MARCH-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6235525man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: MAGE-1 gene

US-08-037-230D-8

Query Match   1.7%; Score 17; DB 4; Length 5674;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    463 ggtcctgggcaccctgg 479
          |||||||||||||||||
Db   3994 GGTCCTGGGCACCCTGG 4010
```

RESULT 28
US-07-853-913-1

; Sequence 1, Application US/0785913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11236 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-853-913-1

Query Match
Best Local Similarity 100.0%;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1.7%; Score 17; DB 1; Length 11236;
Pred. No. 34;

```
Qy     689 ccttcaggaagggggctg 705
           ||||||||||||||||||
Db    3466 CCTTCAGGAAGGGGCTG 3482
```

RESULT 29
US-09-335-409-1/c
; Sequence 1, Application US/0935409
; Patent No. 612029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

```
Qy     268 cggcccagccccagcag 284
           |||||||||||||||||
Db   20332 CGGCCCAGCCCCAGCAG 20316
```

Query Match
Best Local Similarity 100.0%;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1.7%; Score 17; DB 3; Length 68750;
Pred. No. 34;

RESULT 30
US-07-753-110B-12
; Sequence 12, Application US/07753110B
; Patent No. 5436141
; GENERAL INFORMATION:
; APPLICANT: Miyata, Shohei
; APPLICANT: Ohshima, Atsushi
; APPLICANT: Inouye, Sumiko
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: METHOD FOR SYNTHESIZING STABLE
; TITLE OF INVENTION: SINGLE-STRANDED cDNA IN EUKARYOTES BY MEANS OF A BACTERIAL
; TITLE OF INVENTION: RETRON, PRODUCTS AND USES THEREFOR
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/753,110B
; FILING DATE: 30-AUG-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377.5584P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19
; OTHER INFORMATION: /note= "The 2' position of this
; OTHER INFORMATION: nucleotide is linked to the 5' position of
; OTHER INFORMATION: nucleotide number 1 of SEQ ID NO: 11 of this
; OTHER INFORMATION: application."
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 69..76

; OTHER INFORMATION: /note= "This region can hydrogen
; OTHER INFORMATION: bond to nucleotides 156-163 of SEQ ID NO: 11 of
; OTHER INFORMATION: this application."

US-07-753-110B-12

Query Match

Best Local Similarity 87.5%; Score 16; DB 1; Length 76;
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy  977 gtccaagccatcagc  992
        |:||||||||||:|||
Db    5 GUCCCAAGCCAUCAGC  20
```

RESULT 31
US-08-503-730-16
Sequence 16, Application US/08503730
Patent No. 5780269
; GENERAL INFORMATION:
; APPLICANT: Inouye, Sumiko
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: NEW HYBRID MOLECULES
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/503,730
; FILING DATE: 18-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/817,430
; FILING DATE: 06-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377(913).6277P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both

RESULT 32
US-08-507-634-13
Sequence 13, Application US/0850 7634
Patent No. 5845563
; GENERAL INFORMATION:

; APPLICANT: Miyata, Shohei
; APPLICANT: Ohshima, Atsushi
; APPLICANT: Inouye, Sumiko
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: METHOD FOR SYNTHESIZING STABLE
; TITLE OF INVENTION: SINGLE-STRANDED CDNA IN EUKARYOTES BY MEANS OF A BACTERIAL
; TITLE OF INVENTION: RETRON, PRODUCTS AND USES THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,634
; FILING DATE: 25-JUL-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377.6282P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19
; OTHER INFORMATION: /note= "The 2' position of this
; OTHER INFORMATION: nucleotide is linked to the 5' position of
; OTHER INFORMATION: nucleotide number 1 of SEQ ID NO: 12 of this
; OTHER INFORMATION: application."
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 69..76
; OTHER INFORMATION: /note= "This region can hydrogen
; OTHER INFORMATION: bond to nucleotides 156-163 of SEQ ID NO: 12 of
; OTHER INFORMATION: this application."

US-08-507-634-13

Query Match

Best Local Similarity 87.5%; Score 16; DB 2; Length 76;
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy  977 gtccaagccatcagc  992
        |:||||||||||:|||
Db    5 GUCCCAAGCCAUCAGC  20
```

RESULT 33
US-08-905-223-207/c
Sequence 207, Application US/08905223
Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duelert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:

```
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: other
LOCATION: 60..181
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 1..122
OTHER INFORMATION: id AA057454
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 182..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 122..173
OTHER INFORMATION: id AA057454
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 71..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 1..163
OTHER INFORMATION: id C18312
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 182..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 144..195
OTHER INFORMATION: id W69247
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 98..144
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 62..108
OTHER INFORMATION: id W69247
OTHER INFORMATION: est
FEATURE:
```

```
NAME/KEY: other
LOCATION: 34..78
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 1..45
OTHER INFORMATION: id W69247
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 146..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 69..156
OTHER INFORMATION: id H75891
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 76..144
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 95
OTHER INFORMATION: region 1..69
OTHER INFORMATION: id H75891
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: 80..233
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 98
OTHER INFORMATION: region 1..154
OTHER INFORMATION: id HUML11265
OTHER INFORMATION: est
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 104..160
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 7.4
OTHER INFORMATION: seq PMLIRALAQAARA/GP
US-08-905-223-207
Query Match        1.6%;  Score 16;  DB 4;  Length 235;
Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
Matches  16;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   647 ccaggaggtcctgcac 662
         ||||||||||||||||
Db   171 CCAGGAGGTCCTGCAC 156

RESULT  34
US-08-630-822A-97/c
Sequence 97, Application US/08630822A
Patent No. 5840695
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELS, LYNDA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/630,822A
FILING DATE: 11-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-630-822A-97

Qy 289 aagaagtcagtacca 304
        |||||||||||||||
Db 195 AAGAAGTCAGTACCA 180

Query Match   1.6%;   Score 16;   DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 35
US-09-005-069-97/c
Sequence 97, Application US/09005069
Patent No. 5932470
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELS, LYNDA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,069
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,822
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-09-005-069-97

Query Match   1.6%;   Score 16;   DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 aagaagtcagtacca 304
        |||||||||||||||
Db 195 AAGAAGTCAGTACCA 180

RESULT 36
US-08-906-769-104/c
Sequence 104, Application US/0890676769
Patent No. 6077687
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,769
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..251
OTHER INFORMATION:
OTHER INFORMATION: /note= "At pos. bp 4, change A to
R. At pos. aa 2, substitute Xaa."
US-08-906-769-104:

Qy 289 aagaagtcagtacca 304

RESULT 37
US-08-906-616-104/c
Sequence 104, Application US/08906616
Patent No. 6121035
GENERAL INFORMATION:
~ APPLICANT: Grieve, Robert B.
~ APPLICANT: Rushlow, Keith E.
~ APPLICANT: Wu Hunter, Shirley
~ APPLICANT: Frank, Glenn R.
~ APPLICANT: Stiegler, Gary
~ APPLICANT: Gaines, Patrick J.
~ APPLICANT: Silver, Gary
~ TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
~ NUMBER OF SEQUENCES: 190
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: Sheridan Ross P.C.
~ STREET: 1700 Lincoln Street, Suite 3500
~ CITY: Denver
~ STATE: Colorado
~ COUNTRY: USA
~ ZIP: 80203
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: Patentin Release #1.0, Version #1.30
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/906,616
~ FILING DATE: 05-AUG-1997
~ CLASSIFICATION: 536
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Connell, Gary J.
~ REGISTRATION NUMBER: 32,020
~ REFERENCE/DOCKET NUMBER: 2618-25-C2-3
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (303) 863-9700
~ TELEFAX: (303) 863-0223
~ INFORMATION FOR SEQ ID NO: 104:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 252 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ MOLECULE TYPE: cDNA
~ FEATURE:
~ NAME/KEY: CDS
~ LOCATION: 1..251
~ FEATURE:
~ NAME/KEY:
~ LOCATION:
~ OTHER INFORMATION: /note= "At pos. bp 4, change A to
US-08-906-616-104   R. At pos. aa 2, substitute Xaa."

Qy   289 aagaagtcagtacca 304
           ||||||||||||||||
Db   199 AAGAAGTCAGTACCA 184

Query Match   1.6%; Score 16; DB 3; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 38
US-08-817-795-104/c
Sequence 104, Application US/08817795
Patent No. 6139840
GENERAL INFORMATION:
~ APPLICANT: Grieve, Robert B.
~ APPLICANT: Rushlow, Keith E.
~ APPLICANT: Hunter, Shirley Wu

~ APPLICANT: Frank, Glenn R.
~ APPLICANT: Heath, Andrew W.
~ APPLICANT: Yamaka, Miles Yamanaka
~ APPLICANT: Arfsten, Ann
~ APPLICANT: Dale, Beverly
~ APPLICANT: Stiegler, Gary
~ TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
~ TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
~ TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
~ TITLE OF INVENTION: MOLECULES, AND USES THEREOF
~ NUMBER OF SEQUENCES: 119
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: Sheridan Ross & McIntosh
~ STREET: 1700 Lincoln Street, Suite 3500
~ CITY: Denver
~ STATE: Colorado
~ COUNTRY: USA
~ ZIP: 80203
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: Patentin Release #1.0, Version #1.25
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/817,795
~ FILING DATE:
~ CLASSIFICATION:
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: PCT/US95/14442
~ FILING DATE:
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Gary J. Connell
~ REGISTRATION NUMBER: 32,020
~ REFERENCE/DOCKET NUMBER:
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (303) 863-9700
~ TELEFAX: (303) 863-0223
~ INFORMATION FOR SEQ ID NO: 104:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 252 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ MOLECULE TYPE: cDNA
~ FEATURE:
~ NAME/KEY: CDS
~ LOCATION: 1..251
~ FEATURE:
~ NAME/KEY: CDS
~ LOCATION: 1..251
~ FEATURE:
~ NAME/KEY:
~ LOCATION: 2
~   Xaa = any amino acid
US-08-817-795-104

Qy   289 aagaagtcagtacca 304
           ||||||||||||||||
Db   199 AAGAAGTCAGTACCA 184

Query Match   1.6%; Score 16; DB 3; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 39
US-08-639-075A-104/c
Sequence 104, Application US/0863907SA
Patent No. 6150125
GENERAL INFORMATION:
~ APPLICANT: Grieve, Robert B.
~ APPLICANT: Rushlow, Keith E.
~ APPLICANT: Wu Hunter, Shirley
~ APPLICANT: Frank, Glenn R.
~ APPLICANT: Stiegler, Gary
~ APPLICANT: Gaines, Patrick J.

RESULT 40
US-09-012-431-104/c
; Sequence 104, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
     APPLICANT: Grieve, Robert B.
     APPLICANT: Rushlow, Keith E.
                Wu Hunter, Shirley
                Frank, Glenn R.
                Stiegler, Gary
                Gaines, Patrick J.
                Silver, Gary
     TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
                MOLECULES AND USES THEREOF
     NUMBER OF SEQUENCES: 190
     CORRESPONDENCE ADDRESS:
        ADDRESSEE: Sheridan Ross & McIntosh
        STREET: 1700 Lincoln Street, Suite 3500
        CITY: Denver
        STATE: Colorado
        COUNTRY: USA
        ZIP: 80203

Qy   289 aagaagtcagtacca 304
          ||||||||||||||||
Db   199 AAGAAGTCAGTACCA 184

Query Match   1.6%;   Score 16;   DB 3;   Length 252;
Best Local Similarity 100.0%;   Pred. No. 1.1e+02;
Matches 16;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

US-08-639-075A-104

   APPLICANT: Silver, Gary
   TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
   TITLE OF INVENTION: MOLECULES AND USES THEREOF
   NUMBER OF SEQUENCES: 190
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: Sheridan Ross & McIntosh
      STREET: 1700 Lincoln Street, Suite 3500
      CITY: Denver
      STATE: Colorado
      COUNTRY: USA
      ZIP: 80203
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/639,075A
      FILING DATE: 24-APR-1996
      CLASSIFICATION: 424
   INFORMATION FOR SEQ ID NO: 104:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 252 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: cDNA
   FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..251
      OTHER INFORMATION: /note= "At pos. bp 4, change A to
         R. At pos. aa 2, substitute Xaa."

ATTORNEY/AGENT INFORMATION:
   NAME: Connell, Gary J.
   REGISTRATION NUMBER: 32,020
   REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (303) 863-9700
   TELEFAX: (303) 863-0223

RESULT 41
US-09-012-692-104/c
; Sequence 104, Application US/09012692
; Patent No. 6214579
; GENERAL INFORMATION:
     APPLICANT: Grieve, Robert B.
     APPLICANT: Rushlow, Keith E.
     APPLICANT: Wu Hunter, Shirley
     APPLICANT: Frank, Glenn R.
     APPLICANT: Stiegler, Gary
     APPLICANT: Gaines, Patrick J.
     APPLICANT: Silver, Gary
     TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
     TITLE OF INVENTION: MOLECULES AND USES THEREOF
     NUMBER OF SEQUENCES: 190
     CORRESPONDENCE ADDRESS:
        ADDRESSEE: Sheridan Ross & McIntosh
        STREET: 1700 Lincoln Street, Suite 3500
        CITY: Denver
        STATE: Colorado
        COUNTRY: USA
        ZIP: 80203
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/012,692

Qy   289 aagaagtcagtacca 304
          ||||||||||||||||
Db   199 AAGAAGTCAGTACCA 184

Query Match   1.6%;   Score 16;   DB 4;   Length 252;
Best Local Similarity 100.0%;   Pred. No. 1.1e+02;
Matches 16;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

US-09-012-431-104

   R. At pos. aa 2, substitute Xaa."
   SEQUENCE DESCRIPTION: SEQ ID NO: 104:

INFORMATION FOR SEQ ID NO: 104:
   FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..251
      OTHER INFORMATION: /note= "At pos. bp 4, change A to
   MOLECULE TYPE: cDNA
   TOPOLOGY: linear
   STRANDEDNESS: single
   TYPE: nucleic acid
   LENGTH: 252 base pairs
   SEQUENCE CHARACTERISTICS:

APPLICATION NUMBER: US 08/639,075
   FILING DATE: <Unknown>
   CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US/09/012,431
   FILING DATE: 23-Jan-1998
CURRENT APPLICATION DATA:
   SOFTWARE: PatentIn Release #1.0, Version #1.30
   OPERATING SYSTEM: PC-DOS/MS-DOS
   COMPUTER: IBM PC compatible
   MEDIUM TYPE: Floppy disk
COMPUTER READABLE FORM:

ATTORNEY/AGENT INFORMATION:
   NAME: Connell, Gary J.
   REGISTRATION NUMBER: 32,020
   REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (303) 863-9700
   TELEFAX: (303) 863-0223

RESULT 42
US-08-906-613-104/c
Sequence 104, Application US/08906613
Patent No. 6232096
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Silver, Gary
APPLICANT: Gaines, Patrick J.
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; OTHER INFORMATION: /note= "At pos. bp 4, change A to
; US-09-012-692-104
; OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."

Qy 289 aagaaagtcagtaccca 304
        ||||||||||||||||||
Db 199 AAGAAGTCAGTACCA 184

Query Match
Best Local Similarity   1.6%;   Score 16; DB 4; Length 252;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 43
PCT-US95-14442A-104/c
Sequence 104, Application PCT/US9514442A
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.
APPLICANT: Yanaka, Miles Yamanaka
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Stiegler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14442A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid

; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..251
; OTHER INFORMATION: /note= "At pos. bp 4, change A to
; US-08-906-613-104
; OTHER INFORMATION: R. At pos. aa 2, substitute Xaa."

Qy 289 aagaaagtcagtaccca 304
        ||||||||||||||||||
Db 199 AAGAAGTCAGTACCA 184

Query Match
Best Local Similarity   1.6%;   Score 16; DB 4; Length 252;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
;
;
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: cDNA
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: 1..251
;  FEATURE:
;  NAME/KEY:
;  LOCATION: 2
;  Xaa = any amino acid
PCT-US95-14442A-104

Qy    289 aagaagtcagtaccca 304
          ||||||||||||||||
Db    199 AAGAAGTCAGTACCCA 184

Query Match       1.6%;  Score 16;  DB 5;  Length 252;
Best Local Similarity    100.0%;  Pred. No. 1.1e+02;
Matches    16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

RESULT 44
US-09-060-756-254/c
Sequence 254, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 254
LENGTH: 291
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-254


Qy    61 caccagttccgctgt 76
         ||||||||||||||||
Db    83 CACCAGTTCCGCTGT 68

Query Match       1.6%;  Score 16;  DB 4;  Length 291;
Best Local Similarity    100.0%;  Pred. No. 1.1e+02;
Matches    16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

RESULT 45
US-08-334-254-7
Sequence 7, Application US/0834254
Patent No. 5723290
GENERAL INFORMATION:
APPLICANT: James Eberwine, Marc Dichter, Kevin Miyashiro
TITLE OF INVENTION: USE OF NEURITE LOCALIZED mRNAs FOR
TITLE OF INVENTION: MEDICAL DIAGNOSIS AND THERAPEUTICS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
```

```
;  ZIP: 08002
COMPUTER READABLE FORM:
;  MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;  COMPUTER: IBM 486
;  OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
;  SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/334,254
;  FILING DATE: Herewith
;  CLASSIFICATION: 436
PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:
;  FILING DATE:
ATTORNEY/AGENT INFORMATION:
;  NAME: Jane Massey Licata
;  REGISTRATION NUMBER: 32,257
;  REFERENCE/DOCKET NUMBER: PENN-0028
TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (609) 779-2400
;  TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 7:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 489
;  TYPE: NUCLEIC ACID
;  STRANDEDNESS: SINGLE
;  TOPOLOGY: LINEAR
;  ANTI-SENSE: NO
US-08-334-254-7


Qy    693 caggaaggggctgaga 708
          ||||||||||||||||
Db    184 CAGGAAGGGGCTGAGA 199

Query Match       1.6%;  Score 16;  DB 1;  Length 489;
Best Local Similarity    100.0%;  Pred. No. 1.1e+02;
Matches    16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;


Search completed: October 28, 2001, 22:34:10
Job time: 2648 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2001, 21:23:22 ; Search time 1154.11 Seconds
                   (without alignments)
                   8198.791 Million cell updates/sec

Title: SEQ1INS_COPY_1140_2140
Perfect score: 1001
Sequence: 1 tcgcagccacaagattcaaa.........aagccatcagcgtggggatg 1001

Scoring table: OLIGO_NUC
                   Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 472642675O residues

Word size :   8

Total number of hits satisfying chosen parameters: 19872988

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   EST:*

1: gb_est1:*	44: em_esthum10:*	
2: gb_est2:*	45: em_esthum11:*	
3: gb_est3:*	46: em_esthum12:*	
4: gb_est4:*	47: em_esthum13:*	
5: gb_est5:*	48: em_esthum14:*	
6: gb_est6:*	49: em_esthum15:*	
7: gb_est7:*	50: em_esthum16:*	
8: gb_est8:*	51: em_esthum17:*	
9: gb_est9:*	52: em_esthum18:*	
10: gb_est10:*	53: em_esthum19:*	
11: gb_est11:*	54: em_esthum20:*	
12: gb_est12:*	55: em_esthum21:*	
13: gb_est13:*	56: em_esthum22:*	
14: gb_est14:*	57: em_esthum23:*	
15: gb_est15:*	58: em_esthum24:*	
16: gb_est16:*	59: em_esthum25:*	
17: gb_est17:*	60: em_esthum26:*	
18: gb_est18:*	61: em_esthum27:*	
19: gb_est19:*	62: em_esthum28:*	
20: gb_est20:*	63: em_estin1:*	
21: gb_est21:*	64: em_estin2:*	
22: gb_est22:*	65: em_estin3:*	
23: gb_est23:*	66: em_estin4:*	
24: gb_est24:*	67: em_estin5:*	
25: gb_est33:*	68: em_estom1:*	
26: gb_est34:*	69: em_estom2:*	
27: gb_est35:*	70: em_estov1:*	
28: gb_est36:*	71: em_estov2:*	
29: gb_est37:*	72: em_estpl1:*	
30: gb_est38:*	73: em_estpl2:*	
31: gb_est39:*	74: em_estpl3:*	
32: gb_est40:*	75: em_estpl4:*	
33: em_estba:*	76: em_estpl5:*	
34: em_estfun:*	77: em_estpl6:*	
35: em_esthum1:*	78: em_estpl7:*	
36: em_esthum2:*	79: em_estpl8:*	
37: em_esthum3:*	80: em_estpl9:*	
38: em_esthum4:*	81: em_estpl10:*	
39: em_esthum5:*	82: em_estro1:*	
40: em_esthum6:*	83: em_estro2:*	
41: em_esthum7:*	84: em_estro3:*	
42: em_esthum8:*	85: em_estro4:*	
43: em_esthum9:*	86: em_estro5:*	
	87: em_estro6:*	
	88: em_estro7:*	
	89: em_estro8:*	
	90: em_estro9:*	
	91: em_estro10:*	
	92: em_estro11:*	
	93: em_estro12:*	
	94: em_estro13:*	
	95: em_estro14:*	
	96: em_estro15:*	
	97: em_estro16:*	
	98: em_estro17:*	
	99: em_estro18:*	
	100: em_estro19:*	
	101: em_estro20:*	
	102: gb_est25:*	
	103: gb_est26:*	
	104: gb_est27:*	
	105: gb_est28:*	
	106: gb_est29:*	
	107: gb_est30:*	
	108: gb_est31:*	
	109: gb_est32:*	
	110: gb_est41:*	
	111: gb_est42:*	
	112: gb_est43:*	
	113: gb_est44:*	
	114: gb_est45:*	
	115: gb_est46:*	
	116: gb_est47:*	

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
```

```
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

RESULT   1
LOCUS   BE260495   676 bp   mRNA   EST   26-OCT-2000
DEFINITION   601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
             mRNA sequence.
ACCESSION   BE260495
VERSION   BE260495.1 GI:9131807
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 676)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM176 row: d column: 01
             High quality sequence stop: 672.

FEATURES   Location/Qualifiers
     source   1..676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503184"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

BASE COUNT   154 a   207 c   176 g   139 t
ORIGIN

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	509	50.8	676	165	BE260495	BE260495 601150702
2	502	50.1	664	166	BE383336	BE383336 601298249
3	502	50.1	664	166	BE383353	BE383353 601298656
4	470	47.0	692	166	BE382353	BE382353 601298656
5	452	45.2	823	138	BE619259	BE619259 601473130
6	410	41.0	614	169	BG335963	BG335963 602047708
7	399	39.9	960	169	BF797306	BF797306 602256926
8	381	38.1	915	147	BF311926	BF311926 601897767
9	378	37.8	761	140	BE795820	BE795820 601590856
10	371	37.1	884	154	BG519751	BG519751 602578902
11	355	35.5	745	141	BE900936	BE900936 601674206
11	354	35.4	1012	153	BG386348	BG386348 602455550
12	337	33.7	920	153	BG386090	BG386090 602455264
13	330	33.0	992	139	BE747163	BE747163 601577254
14	324	32.4	878	147	BF338283	BF338283 602035507
15	317	31.7	574	165	BE250309	BE250309 600943455
16	313	31.2	762	107	AU124671	AU124671 601899518
17	313	31.2	1497	147	BF315158	BF315158 601899518
18	312	31.2	735	141	BE902696	BE902696 601677393
19	307	30.7	431	5	AA310236	AA310236 EST181085
20	289	28.9	612	166	BE304720	BE304720 601106236
21	289	28.9	307	188	T34216	T34216 EST64346 Hu
22	282	28.2	947	150	BF525432	BF525432 602069517
23	282	28.2	1124	147	BF342802	BF342802 602015077
24	280	28.0	677	139	BE742908	BE742908 601574609
25	262	26.2	482	187	R55841	R55841 yg89d01.r1
26	253	25.3	944	172	BF969043	BF969043 602269966
27	247	24.7	517	104	AI991599	AI991599 ws18c04.x
28	246	24.6	446	10	AA634909	AA634909 ab27h02.r
29	243	24.3	346	122	AW889463	AW889463 RC6-NT002
30	240	24.0	501	5	AA311855	AA311855 EST182568
31	240	24.0	698	140	BE795434	BE795434 601592991
32	230	23.0	938	165	BE260626	BE260626 601146116
33	223	22.3	278	120	AW806551	AW806551 IL0-ST000
34	209	20.9	282	188	T34024	T34024 EST61387 Hu
35	209	20.9	963	154	BG480926	BG480926 602530075
36	206	20.6	472	187	R51138	R51138 yg71c08.r1
37	199	19.9	891	175	BG258460	BG258460 602379938
38	197	19.7	906	152	BG336190	BG336190 602404980
39	195	19.5	961	153	BG387104	BG387104 602455813
40	195	19.5	394	6	AA362268	AA362268 EST52407
41	185	18.5	891	174	BG116283	BG116283 602318546
42	185	18.5	547	168	BF686235	BF686235 602143687
43	182	18.2	688	140	BE794311	BE794311 601591442
44	179	17.9	852	146	BF240253	BF240253 601905776
45	174	17.4	441	157	H03317	H03317 yj47e10.r1

ALIGNMENTS

Query Match   50.8%;   Score 509; DB 165;   Length 676;
Best Local Similarity   99.8%;   Pred. No. 1.1e-259;
Matches   629;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   1;

```
Qy      1  tcgcagccacaagattcaaaccagctcaacctcatccaccggacatcttcccctgct   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     48  TCGCAGCCACAAGATTCAAACCAGCTCAACCTCATCCACCGGACATCTTCCCCTGCT  107

Qy     61  caccagtttccgctgtaagaaggaggccccaccctcagtgtgccaatggttcagggtga  120
           ||||||||||||||||||||||||||| |||||| ||||||||||| ||||||||||||
Db    108  CACCAGTTTCCGCTGTAAGAAGGAGGGCCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGA  167

Qy    121  atgcctccctcaagtaccagctccgtccaggaggagggatggcagatgccattattac  180
           |||||||||||||||||||||||||||||||||||||| || |||| ||||||||||||
Db    168  ATGCCTCCCTCAAGTACCAGCTCCGTCCAGGAGGAGTGCGGAGGGATGCCATTATTAC  227

Qy    181  ttgcaatcctgaggaattcatagttgaggcgctgcagcttccaacttccagcagagcgt  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    228  TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCAACTTCCAGCAGAGCGT  287

Qy    241  gcagagagtacaaggaggagtgcgcaggacggacccagccagcagagaaagaagtcagta  300
           |||| || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    288  GCAGGAGTACAGGAGGAGTGCGCAGGACGGACCCAGCCAGCAGAGAAAGAAGTCAGTA  347

Qy    301  cccaggaaatacatccttccttggaacaggggtctgccatcccgatgaagaattcgaaatgtcag  360
           ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    348  CCCAGAAAATCATCCTTCCTTGGAACAGGGGTCTGCCATCCCGATGAAGAATTCGAAATGTCAG  407

Qy    361  tgccacacttgtcaacataagcccgacacagtgtctgctactggactgtggtgaggcac  420
           |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db    408  TGCCACACTTGTCAACATAAGCCCGACACACGTCTCTGCTACTGGACTGTGGTGAGGCAC  467

Qy    421  atttgggcagctgtgccgtcattacggagaccaagtgggacaggggtcctggcacctggc  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    468  ATTTGGGCAGCTGTGCCGTCATTACGGAGACCAAGTGGGACAGGGGTCCTGGCACCTGGC  527

Qy    481  tgctgtgtttgtgtcccaactgggacgcagcagccacacacgggcttgccaagtatcttgc  540
           ||||||||||||||||||| |||   ||||||| ||| |||||||||||||||||||||||
Db    528  TGCTGTGTTTGTGTCCCACCT-GCAGCAGATCACCACACGGGCTTGCCAAGTATCTTGC  586

Qy    541  tgcagagagaacgcgccttggcatcttgggaagccttcaccctttgctggtggttg  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    587  TGCAGAGAGAACGCGCCTTGGCATCTTGGGAAGCCTTCACCCTTTGCTGGTGGTTG  646
```

```
Qy    601 ccccaaccagctcaaagcttggctccagc 630
          ||||||||||||||||||||||||||||||
Db    647 CCCCCACCAGCTCAAAGCCTGGCTCCAGC 676
```

RESULT   2

```
BE383336
LOCUS       BE383336        664 bp    mRNA                    EST       21-JUL-2000
DEFINITION  6012982491F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
            mRNA sequence.
ACCESSION   BE383336
VERSION     BE383336.1  GI:9328701
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 664)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM313 row: a column: 13
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1..664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3628308"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT    154 a    199 c    174 g    137 t
ORIGIN

Query Match         50.1%;  Score 502;  DB 166;  Length 664;
Best Local Similarity  100.0%;  Pred. No. 6e-256;
Matches  502;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 tcgcagcacaagatcaaacctcaaacctcatcacaccggacacatcttccccgct 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    48 TCGCAGCACAAGATCAAAACCTCAACCTCATCCACACCGGACATCTTCCCCCGCT 107

Qy    61 caccagtttccgctgtaagaaggaggcccacacctcagtgttgcccatgttcaggttga 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   108 CACCAGTTTCCCGCTGTAAGAAGGAGGGCCCCACACCTCAGTGTGCCCATGGTTCAGGTTGA 167

Qy   121 atgcctcctcaagtaccagctccggtcccaggaggaggagggatgccattattac 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   168 ATGCCTCCTCAAGTACCAGCTCCGGTCCCAGGAGGAGGTGGCAGAGGGATGCCATTATTAC 227

Qy   181 ttgcaatcctggaggattcatagttgaaggcgctgcagggttcccaacttccagcagagcgt 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   228 TTGCAATCCTGGAGGAATTCATAGTTGAGGCGCTGCAGGGCCCTGCAGCTTCCCAACTTCAGCAGAGCGT 287

Qy   241 gcaggagtacaggaggagtcgcaggagacggcccagcccagcagagagaaagaagtcagta 300
         |||||||||||||||||||||
```

RESULT   3

```
BE382353
LOCUS       BE382353        692 bp    mRNA                    EST       21-JUL-2000
DEFINITION  6012908659F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
            mRNA sequence.
ACCESSION   BE382353
VERSION     BE382353.1  GI:9327718
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 692)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM314 row: o column: 13
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1..692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629028"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT    152 a    193 c    197 g    150 t
ORIGIN

Query Match         47.0%;  Score 470;  DB 166;  Length 692;
Best Local Similarity  99.8%;  Pred. No. 6.7e-239;
Matches  590;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy   336 tccggatgaagattcaagtgcccacacttgtcaaacataagccccccgacacgtctc 395
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TCCGGATCGAAGATTCGAAATGTCAGTGCCCACACTTGTCAACACATAAGCCCCGACACGTCTC 60
```

Db   288 GCAGGAGTACAGGAGGAGTGCCGAGGACGGACGGGCCCCAGCACCCAGCAGAGAAAGAAGTCAGTA 347

```
Qy   301 cccaggaaatcacatcttccttggaacaggtgtctgccatcccgatggaagttcgaaatgtcag 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   348 CCCAGAAATCATCTTCCTTGGAACAGGGTGTCTGCCATCCCGATGAAGATTCGAAATGTCAG 407

Qy   361 tgcacacttgtcaaacataagccccgacacagtctctgctacttggactgtgttgaggcac 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   408 TGCCACACTTGTCAACACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCAC 467

Qy   421 atttgggcagctgtgccgtcattacggagacaccaggttggacagggtcctgggcaccctggc 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   468 ATTTGGGCAGCTGTGCCGTCATTACGGAGACACCAGGTGGACAGGGTCCTGGGCACCCTGGC 527

Qy   481 tgctgtgttgtctgtcccacctg 502
         |||||||||||||||||||||
Db   528 TGCTGTGTTGTTTGTGTCCCACCTG 549
```

```
Qy    396  tgctactggactgtgtgtgaggggcacatttgggggcacgtgtgcgtcattacggagaccagg   455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  TGCTACTGGACTGTGGTGGAGGGCACATTTGCGCAGTGTGCCGTCATTACGGAGACCAGG    120

Qy    456  tggacagggtcctggcacccctggctgctgtgttttgtgtcccacctggcacgagatcac    515
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TGGACAGGGTCCTGGCACCCTGGCCTGCTGTGTTTGTGTCCCACCT-GCACGCAGATCAC    179

Qy    516  cacacgggcttgtccagatatctgtcaggagagaacgcgccttgggaatcttctgggaaag   575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180  CACACGGGCTTGTCCAGATATCTGTTCCAGAGAACGCGCCGCATCTTTGGGAAAG        239

Qy    576  ccgcttcaccctttgctgttgggttgcccccaaccagctccaaagcctggctccagcagtac   635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240  CCGCTTCACCCTTTGCTGTTGGGTTGCCCCCAACCAGCTCCAAAGCCTGGCTCCAGCAGTAC   299

Qy    636  cacaaccagtgccagagggagtcctgaccaccatcagtatgatgatccttgccaaatgcctcag   695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300  CACAACCAGTGCCAGAGGTCCTGCCACCACCAGTATGATTCCTGGCCAAATGCCTCAG    359

Qy    696  gaagggggctgagatctccagtcgcagtggaaagattgatcaagttcgcgtgttgcgaaca   755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360  GAAGGGGCTGAGATCTCCAGTCGCGAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACA   419

Qy    756  tgtgattttgaagagtttcagacctgtctgtgtgcggcactgcaagcatgcgtttgctgttgctgt   815
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420  TGTGATTTTGGAAGAGTTTCAGACCTGTCTGTGTGCGGCACTGCAAGCATGCGTTTGGCTGT   479

Qy    816  gcgctggtgcacacctctggcgaaagtgctgaaaagtggtctctattccgggggacac   875
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480  GCGCTGGTGCACACCTCTGGCGAAAGTGGTCTATTCCGGGGGACAC    539

Qy    876  gctctggttccggatggggaaagaagatgccaccctccctgatacatgaagccacc   926
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540  GCTCTGGTTCCGGATGGGGAAAGAAGATGCCACCCTCCCTGATACATGAAGCCACC   590
```

```
RESULT    4
LOCUS         BE619259       823 bp         mRNA                    EST       20-OCT-2000
DEFINITION    601473130F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5',
              mRNA sequence.
ACCESSION     BE619259
VERSION       BE619259.1   GI:9890197
KEYWORDS      EST.
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 823)
  AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
  TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL     Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: DCTD/DTP/Gazdar
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM9636 row: k  column: 08
              High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1..823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3876223"
                     /clone_lib="NIH_MGC_68"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
```

```
BASE COUNT     185 a     219 c     259 g     160 t
ORIGIN
               Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
               Average insert size 1.8 kb.  Library constructed by Life
               Technologies."
```

```
Query Match            45.2%;   Score 452;  DB 138;   Length 823;
Best Local Similarity  100.0%;  Pred. No. 2.6e-229;
Matches    452;  Conservative    0;  Mismatches    0;   Indels    0;   Gaps    0;

Qy    503  gcacgcagatcacacacacgggcttgtccagatatctgtctgcaggagagaacgcgccttggc   562
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     88  GCACGCAGATCACACACACGGGCTTGCCAGATATCTGTCCAGAGAGAACGCGCCTTGGC    147

Qy    563  atctttgggaaagacgcgtccagtgggttgctgttggtgttgcccccaaccagctcaaagcctg   622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    148  ATCTTTGGGAAAGACGCGTCCAGTGGGTTGTTGGTGTTGCCCCCAACCAGCTCAAAGCCTG   207

Qy    623  gctccagcagtaccacaaccagtgccagagggagtcctgccaccacatcagtatgatgatcctgc   682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    208  GCTCCAGCAGTACCACAACCAGTGCCAGAGGGAGTCCTGCCACCACATCAGTATGATTCCTGC   267

Qy    683  caaatgcctcagcagtgaagatggatcttcagcagtagaagttccgcggcactgcaagcatgcg   742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268  CAAATGCCTCAGCAGTGAAGGGCTGAGATCTCCAGTCGCGAGTGGAAAGATTGATCAGTTC   327

Qy    743  gctgttgcgaacatgtgattttgaagagtttcagacctgtctgtgtgcggcactgcaagca   802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    328  GCTGTTGCGAACATGTGATTTTGGAAGAGTTTCAGACCTGTCTGTGTGCGGCACTGCAAGCA   387

Qy    803  tgcgtttggctgtgcgctggtgcacacctctggctgggaaagtggtctattccgggggacac   862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    388  TGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCGAAAGTGGTCTATTCCGGGGGACAC    447

Qy    863  catgccctcgcgaggcctctcgtctgctggtgtggccacccctccctgatacatgaagc   922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    448  CATGCCCTCGCGAGGCCTCTCGTCTGGCTGGTGCACACCCTCCCTGATACATGAAGC   507
```

```
RESULT    5
LOCUS         BG335963       614 bp         mRNA                    EST       27-FEB-2001
DEFINITION    602404708F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542553 5',
              mRNA sequence.
ACCESSION     BG335963
VERSION       BG335963.1   GI:13142401
KEYWORDS      EST.
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 614)
  AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
  TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL     Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1222 row: o  column: 02
              High quality sequence stop: 614.
FEATURES             Location/Qualifiers
     source          1..614
```

```
Qy    923  caccctggaagatggtttggaagagaagcag   954
           |||||||||||||||||||||||||||||||
Db    508  CACCCTGGAAGATGGTTTGGAAGAGAAGCAG   539
```

AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
REFERENCE   1 (bases 1 to 960)
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM   Homo sapiens
SOURCE   human.
KEYWORDS   EST.
VERSION   BF797306.1   GI:12102360
ACCESSION   BF797306
                mRNA sequence.
DEFINITION   602256926F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4340286 5',
LOCUS   BF797306   960 bp   mRNA   EST   12-JAN-2001
BF79306
RESULT   6

Qy   846 gtctattccgggagacaccatgccctgccgaggctctcggtccggatgggga 896
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GTCTATTCCGGGAGACCATGCCCTGCCGAGGCTCTCGGTCCGGATGGGGAAA 531

Qy   786 gtgcggcacctgcaagcaagtcggtttggctgtcgtgtgcacacctctggaaagtg 845
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTGCGGCACCTGCAAGCATGCCGTTTGGCTGTCGTGCACACCTCTGGGTCTGGAAAGTG 480

Qy   726 gaaagattgatcagttcgctgttgcgaacatgtgattgattttggaaagagttcagacctgtctg 785
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAAAGAATTGATCAGTTCGCTGTTGCGAACATGTGATTGATTGGAAGAGTTTCAGACCTGTCTG 420

Qy   666 atcagtatgattcctgccaaatgcttcaggaaaggggctaagatctccagtcctgcagtg 725
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAGTG 360

Qy   606 aaccagctcaaagctggctccagcagtaccacacaccagtgccagaggtcctgcaccac 665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AACCAGCTCAAAGCTGGCTCCAGCAGTACCACACCAGTGCCAGAGGTCCTGCACCAC 300

Qy   546 agagaagacgcgcttggcatcttggaaagcgcgctcacccttgctgttggttgccccc 605
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGAGAACGCGGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCC 240

Qy   486 tgtttggtgtcccacctggcacgcagatcaacacaacagggcttgccaagtatctgtcag 545
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TGTTTGGTGTCCCACCT-GCACGCAGAATCACCACACGGGCTTGCCAAGTATCTTGCTGCAG 180

Qy   426 ggcagctgtgccgtcattacggacgaccaggtggacaggtcctgggcaccctggctgctg 485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GGCAGCTGTGCCGTCATTACCGGACAGGGTGCGACAGGGTCCTGGGCACCCTGGCTGCTG 121

Qy   366 cacttgtcacataagccccgacacagtgtctctgctactggtgtgtgagggcacattg 425
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 CACTTGTCACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGTGGTTGGAGGGCACATTTG 61

Matches   530; Conservative   0; Mismatches   0; Indels   1; Gaps   1;
Best Local Similarity   99.8%; Pred. No. 5.9e-207;
Query Match   Score 410; DB 152; Length 614;

ORIGIN
BASE COUNT   135 a   169 c   178 g   132 t

/note="Organ: Placenta; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
                /lab_host="DH10B (phage-resistant)"
                /tissue_type="choriocarcinoma"
                /clone_lib="NIH_MGC_21"
                /clone="IMAGE:4542553"
                /db_xref="taxon:9606"
                /organism="Homo sapiens"

```
Qy    823  tgcacacctctggctgaaagtggtctattccggggaccacatgccctgcgtgcggagggctctgg  882
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600  TGCACACCTCTGGCTGTGGAAAGTGGTCTATTCCGGGGACCACCATGCCCTGCGTGCGGAGGGCTCTGG  659

Qy    883  tccggatggggaa  895
                |||||||||||||
Db    660  TCCGGATGGGGAA  672
```

```
RESULT   7

LOCUS           BF311926         915 bp    mRNA                          EST    21-NOV-2000
DEFINITION      601897767F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126864 5',
                mRNA sequence.
ACCESSION       BF311926
VERSION         BF311926.1   GI:11259697
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 915)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: ATCC
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
                Plate: LLCM1017 row: f column: 17
                High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1..915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4126864"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA Synthesis Kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT     206 a    262 c    288 g    159 t
ORIGIN

Query Match         38.1%;  Score 381; DB 147;  Length 915;
Best Local Similarity   99.8%;
Matches   501;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy      1  tcgcagcccacaagattcaaacccagctcaacctcatccaccggacatctccccctgct  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68  TCGCAGCCCACAAGATTCAAACCCAGCTCAACCTCATCCACCGGACCATCTTCCCCCTGCT  127

Qy     61  caccagttccgctgtaagaaggaggagggccccaacctcagtgttccatggttcagggtga  120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    128  CACCAGTTCCGCTGTAAGAAGGAGGAGGGCCCCAACCTCAGTGTTGCCATGGTTCAGGGTGA  187

Qy    121  atgcctcctcaaagtaccagctccgtccccaggagaggggatgccaattattac  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188  ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGAGGGGATCCCCATTATTAC  247

Qy    181  ttgcaatcctgaggaattccatagttgaggcgctgcagcttcccaacttccagcagagcgt  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248  TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT  307
```

```
Db    248  TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT  307
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241  gcaggagtacaggagagtgcggcagacgcgacgcgccccagccacagaggaaaagtcagta  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308  GCAGGAGTACAGGAGAGTGCCCAGCCAGGACGGG-CCAGCCCCAGCAGAGAAAAGAAGTCAGTA  366

Qy    301  cccagaaaatcaatcttccttggaacaaggtctgccatccatccgatgaagatccgaaatgtcag  360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367  CCCAGAAAATCAATCTTCCTTTGGAACACAGGGTCTGCCATCCGATGAAGATTCGAAATGTCAG  426

Qy    421  atttgggcagctgtgccgtcattacggagaccaaggtggacagggttcctgggcaacctggc  480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427  TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTCGCTTACTTGGTGGTGGAGGGCCAC  486

Qy    481  tgctgtgtttgtgtcccacctg  502
                ||||||||||||||||||||||
Db    547  TGCTGTGTTTGTGTCCCACCTG  568
```

```
RESULT   8

LOCUS           BE795820         761 bp    mRNA                          EST    20-SEP-2000
DEFINITION      601590856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
                mRNA sequence.
ACCESSION       BE795820
VERSION         BE795820.1   GI:10217018
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 761)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: DCTD/DTP
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
                Plate: LLCM802 row: p column: 14
                High quality sequence stop: 761.
FEATURES             Location/Qualifiers
     source          1..761
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3945085"
                     /clone_lib="NIH_MGC_7"
                     /cell_line="MGC3"
                     /tissue_type="small cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT     172 a    206 c    219 g    164 t
ORIGIN

Query Match         37.8%;  Score 378; DB 140;  Length 761;
Best Local Similarity   99.8%;                  Pred. No. 6.6e-190;
```

Matches 498; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy    503 gcacgcagatcaaccacacggcgttgccaagtatcttgctgcagagagaacgcgccttggc 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    163 GCACGCAGATCAACCACACGGCGTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 222

Qy    563 atccttgggaaagccgcttcaccctttgctggtggttgcccccaacgctcaaagcctg 622
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    223 ATCCTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACGCTCAAAGCCCTG 282

Qy    623 gctccagcagtaccacaacacagtgccagaggaggtcctgcacacatcagtatgattcctgc 682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    283 GCTCCAGCAGTACCACAACACAGTGCCAGAGGAGGTCCTGCACACATCAGTATGATTCCTGC 342

Qy    683 caaatgccttcaggaaggggcgtgagatctccagtcctgcagtggaagattgatcagttc 742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    343 CAAATGCCTTCAGGAAGGGGCGTGAGATCTCCAGTCCTGCAGTGGAAAGATTGATCAGTTC 402

Qy    743 gctggtcgaacatgtgatttggaagagagtttcagaacctgtctgctgcgcactgcaagca 802
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    403 GCTGGTTCGCGAACATGTGA-TTGGAAGAGTTTCAGAACCTGTCTGCTGCGCACTGCAAGCA 461

Qy    803 tgcgtttggctgtcgctggtgtgtacaacctctggctggaaagtgggtctcattccgggacac 862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    462 TGCGTTTGGCTGTCGTCGCCGTGGTCACACCTCTGGAAAGTGGGTCTCATTCCGGGACAC 521

Qy    863 catgtcgcctgcgagtctctgtctccggatgtgggagaatgtcccacocctcctgatacatgaagc 922
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    522 CATGTCCCTGCCGAGCGCTGGGTCGTCGGAGGGAAGATGCCACCTCCTGATTCACATGAAGC 581

Qy    923 cacccctggaagatgtttggaagagaacgtgaaaagacaacaacaacagtcccca 982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    582 CACCCCTGGAAGATGGTTTGGAAGAGAAGAGAAGACAACAGCACAACGTCCCA 641

Qy    983 agccatcaagcgtggggatg 1001
          ||||||||||||||||||||
Db    642 AGCCATCAAGCGTGGGGATG 660
```

/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT   199 a   251 c   249 g   185 t
ORIGIN

Query Match   37.1%; Score 371; DB 154; Length 884;
Best Local Similarity   99.8%; Pred. No. 3.6e-186;
Matches 491; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy    176 attacttgcaatcctgagaaattcaatagttgaggcgctgcaagcttcccaacttccaagcag 235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 ATTACTTGCAATCCTGAGGAAATTCATATAGTTGAGGCGCGTGCAAGCTTCCCAACTTCCAGCAG 332

Qy    236 agcgtgcaggagttacaggaggaggtgcgcaggaacgcgcccagcccagcagagagaaagaagt 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 AGCGTGCAGGAGTTACAGGAGGAGGTGCGCAGGAACGCGCCCAGCCCAGCAGAGAAAAGAAGT 392

Qy    296 cagtacccagaaaatcatcttccttgggaacaggtctgcaatcccgatgaaagtcgaaat 355
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGTACCCAGAAAATCATCTTCCTTGGGAACAGGGTCTGCCATCCCGATGAAAGATTCGAAAT 452

Qy    356 gtcagtgccacacttgtcaacataagccccgacacgtctctgctactggacttgtgttgag 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 GTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAG 512

Qy    416 ggcacattgggcagctgtgtccgtcattacggagaaccaggtggacaaggtcctgggcacc 475
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GGCACATTTGGGCAGCTGTGTCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACC 572

Qy    476 ctggctgctgttgtttgtgtcccaacctgcacgcagatcaccaccacacaacgggcgcttgcaagtat 535
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTGGCTGCTGTGTTGTTTGTGTCCCAACCT-GCACGCAGATCACCACCACACGGGCGCTTGCCAAGTAT 631

Qy    536 cttgctgcagagagaacgcgccttggcatcttgggaaagccgcttcacccttttgctgggt 595
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    632 CTTGCTGCAGAGAGAACGCGCCTTGGCATCTTGGGAAAGCCGCTTCACCCTTTGCTGGT 691

Qy    596 ggttgcccccaacccagctcaaagcctggctggtccagcagtaccacaacacagtgccaggaggt 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    692 GGTTGCCCCCAACCCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACACAGTGCCAGGAGGT 751

Qy    656 cctgcacacat 667
          |||||||||||
Db    752 CCTGCACACAT 763
```

Left column (RESULT 11)

```
FEATURES
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Incyte Genomics, Inc.
        cDNA Library distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
        Plate: LLCM834 row: k column: 01
        High quality sequence stop: 734.
    source          Location/Qualifiers
        1..745
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3957240"
        /clone_lib="NIH_MGC_21"
        /tissue_type="choriocarcinoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCACGAG(G). Size-selected >500bp
        for average insert size 1.8kb. Library constructed by
        Ling Hong in the laboratory of Gerald M. Rubin (University
        of California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    165 a    206 c    215 g    159 t
ORIGIN
```

```
Query Match      35.5%;  Best Local Similarity   100.0%;
Matches  355;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Score 355;  DB 141;  Length 745;

Qy   503  gcacgcaagatcaccacaacggccttgccaagtatcttgctgcagagagaggaacggccttggc  562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117  GCACGCAGATCACCACACGGGCGTTGCCAAGTATCTTGCTGCAGAGAACGCCCTTGGC  176

Qy   563  atcttgggaaagccgcttcaccctttgctgctgtgttgccccaacccagctcaaagcctg  622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177  ATCTTTGGGAAAGCCCGTTCACCCTTTGCTCTGGTTGCCCCAACCCAGCTCAAAGCCTG  236

Qy   623  gctccagcagtaccacacaaccagtgccaggagggtgtcctgccaccacatcagtatgatgatttcctgc  682
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237  GCTCCAGCAGTACCACACAACCAGTGCCAGGAGGGTCTCGACCACCACATCAGTATGATTCCTGC  296

Qy   683  caaatgccttcaggaagggggtgagatcctccagtcctgcagtggaaagattgatcagttc  742
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297  CAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCGGTGGAAAGATTGATCAGTTC  356

Qy   743  gctgtgttgccgaacatgtgattggaagagtttcagaccctgtctgtgcggcactgcaagca  802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357  GCTGTGTTGCCGAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCA  416

Qy   803  tgcgtttgggcgtggtcgcgctggtcacacctctggctgaaagtggtctcattccggg  857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417  TGCGTTTGGGCTGTGGCGCTGGTCGCACACCTCTGGCTGAAAGTGGTCTGATTCCGGG  471
```

```
RESULT 11
BG386348
LOCUS       BG386348      1012 bp     mRNA      EST      12-MAR-2001
DEFINITION  602455550F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583749 5',
            mRNA sequence.
ACCESSION   BG386348
VERSION     BG386348.1  GI:13279794
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

Right column (RESULT 12)

```
FEATURES
        Tissue Procurement: ATCC
        cDNA Library Preparation: Ling Hong/Rubin Laboratory
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: NIH Intramural Sequencing Center
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM1308 row: c column: 14
        High quality sequence stop: 675.
    source          Location/Qualifiers
        1..1012
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4583749"
        /clone_lib="NIH_MGC_15"
        /tissue_type="adenocarcinoma cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
        EcoRI; cDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCACGAG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    246 a    273 c    281 g    210 t    2 others
ORIGIN
```

```
Query Match      35.4%;  Best Local Similarity   100.0%;
Matches  354;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Score 354;  DB 153;  Length 1012;

Qy     1  tcgcagccacaagattcaaacccagctcaatccaccccggacatcttccccctgct  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247  TCGCAGCCACAAGATTCAAACCCAGCTCAATCCACCCGGACATCTTCCCCCTGCT  306

Qy    61  caccagttccgctgtaagaaggagggggcccacaccctcagtgtgtgccatggttcagggtga  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307  CACCAGTTCCGCTGTAAGAAGGAGGGGCCCACACCCTCAGTGTGTGCCATGGTTCAGGGTGA  366

Qy   121  atgcctcctcaaagtaccacagtctccgtcccaggaggagaggtggcagaggaatgccattattac  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367  ATGCCTCCTCAAAGTACCACAGTCTCCGTCCCAGGAGGAGTGGCAGAGGAGGATCCATTATTAC  426

Qy   181  ttgcaatcctgaggaattcataghtgaggcgctgcagcttcccaacttccagcagagagcgt  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427  TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT  486

Qy   241  gcaggagtacagggaggagtgcgcaggagacggccgccccagcaagacagaaaagaagtcaagta  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   487  GCAGGAGTACAGGGAGGAGTGCGCAGGAGACGGCCGCCCCAGCAAGACAGAGAAAGAAGTCAAGTA  546

Qy   301  cccagaaaatcatcttcccttggaaacaggggtctgccatcccgatgaaagattcgaaa  354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   547  CCCAGAAAATCATCTTCCTTGGAAACAGGGGTCTGCCATCCCGATGAAAGATTCGAAA  600
```

```
RESULT 12
BG386090
LOCUS       BG386090       920 bp     mRNA      EST      12-MAR-2001
DEFINITION  602455264F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583679 5',
            mRNA sequence.
ACCESSION   BG386090
VERSION     BG386090.1  GI:13279536
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

Left column

```
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1307 row: p column: 16
            High quality sequence stop: 736.
FEATURES
     source          1..920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583679"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      196 a    261 c    272 g    190 t    1 others
ORIGIN

Query Match   Best Local Similarity   99.8%;
Matches  457; Conservative  0; Mismatches  0; Indels  0; Gaps  1;
              Score 337; DB 153;   Length 920;   Pred. No. 4.6e-168;

Qy    1 tcgcagccacacaagattcaaaccaagtcaacctcatccaaccggacaatcttccccctgct 60
        ||||||||||||| ||||||||||| |||||||||||| |||||||||||||| |||||||||||
Db  245 TCGCAGCCACAAGATTCAAACCCAGTCAACCTCACCTCATCCAACCGGACACATCTTCCCCTGCT 304

Qy   61 caccagttctccgctgtaagaagaggaggcccccaccccctcagttgtgccatggttcagggtga 120
        ||||||||||||||||||||||||||||||||||||     ||||| ||||||||||||||||||
Db  305 CACCAGTTCTCCGCTGTAAGAAGAGGAGGCCCCCAGTGTGTGCCCATGGTTCAGGGTGA 364

Qy  121 atgcctcctcaagtaccagctccgtcccagcaggagggagtggcagaggggatgccattattac 180
        ||||||||||||||||||||||||||||||| |||||||| |||||   |||||||||||||||
Db  365 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGGGAGGGATGCCATTATTAC 424

Qy  181 ttgtcaatcctgaggaattcatagttgaggcgctgcaagcttcccaacttccagcagagcgt 240
        |||| ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db  425 TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT 484

Qy  241 gcaggagtacagcaggagagtcgcgcaggacgcccagcagagaaaagaagtcagta 300
        |||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  485 GCAGGAGTACAGCAGGAGGACGCGCCAGGACGCCCAGCAGAGAAAAGAAGTCAGTA 544

Qy  301 cccagaaatacatcttccttggaacaggggtctgccatcccgatgaagattcgaaatgtcag 360
        ||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  545 CCCAGAAATCATCTTCCTTGGAACAGGGTCTG-CATCCCGATGAAGATTCGAAATGTCAG 603

Qy  361 tgccacacttgtcaacataagccccgacacgtctctgtactgtggtgaggcac 420
        |||||||||||||||||||||||||||||||||||||||||| ||| |||||||
Db  604 TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGTACTGGTGTGGTGAGGGCAC 663

Qy  421 atttgggcagctgtgccgtcattacggagaccaggtgg 458
        |||||||||||||||||||||||||||||||||||||||
Db  664 ATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGG 701
```

```
RESULT 13
LOCUS       BE747163      992 bp    mRNA         EST       15-SEP-2000
DEFINITION  801577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
```

Right column

```
ACCESSION   BE747163
VERSION     BE747163.1  GI:10161155
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 992)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
            Plate: LLCM524 row: l column: 14
            High quality sequence stop: 781.
FEATURES
     source          1..992
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3838237"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) and Superscript II RT (Life Technologies)."
BASE COUNT      225 a    287 c    313 g    167 t
ORIGIN

Query Match   Best Local Similarity   99.6%;
Matches  500; Conservative  0; Mismatches  0; Indels  1; Gaps  1;
              Score 330; DB 139;   Length 992;   Pred. No. 2.5e-164;

Qy    1 tcgcagccacaagattcaaaccaagtcaacctcatccaaccggacaatcttccccctgct 60
        |||||||||||||||||||||||| |||||||||||| |||||||||||||| |||||||||||
Db  148 TCGCAGCCACAAGATTCAAACCCAGTCAACCTCACCTCATCCAACCGGACACATCTTCCCCTGCT 207

Qy   61 caccagttctccgctgtaagaagaggaggcccccaccccctcagttgtgccatggttcagggtga 120
        ||||||||||||||||||||||||||||||||||||||| | |||||  |||||||||||||||||
Db  208 CACCAGTTCTCCGCTGTAAGAAGAGGAGGCCCCCACCCTCAGTGTCCCCARGGTTCAGGGTGA 267

Qy  121 atgcctcctcaagtaccagctccgtcccagcaggagggagtggcagaggggatgccattattac 180
        ||||||||||||||||||||||||||||||| |||||||| |||||   |||||||||||||||
Db  268 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGGGAGGGATGCCATTATTAC 327

Qy  181 ttgtcaatcctgaggaattcatagttgaggcgctgcaagcttcccaacttccagcagagcgt 240
        |||| ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db  328 TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGT 387

Qy  241 gcaggagtacagcaggagagtcgcgcaggacgcccagcagagaaaagaagtcagta 300
        ||||||||||||||||| |||||||||||||    ||||| ||||||||||||||||
Db  388 GCAGGAGTACAGCAGGAGGACGCGCCAGGACGG--CCAGCCCCAGCAGAGAAAAGAAGTCAGTA 446

Qy  301 cccagaaatacatcttccttggaacaggggtctgccatcccgatgaagattcgaaatgtcag 360
        ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  447 CCCAGAAATCATCTTCCTTGGAACAGGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAG 506

Qy  361 tgccacacttgtcaacataagccccgacacgtctctgtactgtggtgaggcac 420
        |||||||||||||||||||||||||||||||||||||||||| ||| |||||||
Db  507 TGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGTACTGGTGTGGTGAGGGCAC 566
```

RESULT 14

```
Qy    421 atttgggcaagctgtcgcgtcattacggagacaaggtggacaggggtcctgggcacccggc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 ATTTGGGCAGCTGTCGCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGC 626

Qy    481 tgctgtgtttgttgtcccacctg 502
          |||||||||||||||||||||||
Db    627 TGCTGTGTTTGTTGTCCCACCTG 648
```

LOCUS   BF338283   878 bp   mRNA   EST   22-NOV-2000
DEFINITION   602035507F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183488
             5', mRNA sequence.
ACCESSION   BF338283
VERSION   BF338283.1 GI:11284683
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 878)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9499 row: n column: 01
             High quality sequence stop: 631.
FEATURES   Location/Qualifiers
     source   1..878
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4183488"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.57 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   208 a   250 c   230 g   190 t
ORIGIN

Query Match   32.4%; Score 324; DB 147; Length 878;
Best Local Similarity   99.8%;
Matches 444; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

```
Qy      1 tcgcagccacaagatcaaacccaagctcaacctcatccaccggacatcttccccctgct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175 TCGCAGCCACAAGATCAAACCCAAGCTCAACCTCATCCACCGGACAATCTTCCCCTGCT 234

Qy     61 caccagttccgctgtaagaggaggagggccccaccctcagtgtgtccccatgggttca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 CACCAGTTTCCGCTGTAAGAGGAGGAGGGCCCCACCCTCAGTGTGCCCCATGGGTTCAGGGTGA 294

Qy    121 atgcctcctcaagtaccacgtccgtccaggagggagtggagccattattac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    295 ATGCCTCCTCAAGTACCACGTCCGTCCCAGGAGGGAGGGGATGCCATTATTAC 354

Qy    181 ttgcaatcctgaggaattcataagttgagcgcgctgcagcttcccaactccagcagagacgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    355 TTGCAATCCTGAGGAATTCATTAGTTGAGGCGCTGCAGCGTTCCCAACTTCCAGCAGAGAGGT 414
```

RESULT 15

```
Qy    241 gcaggagtacaggaggagtgcgcaggacggcaggacgggcccagcccagcccagcagagaaagtcagta 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    415 GCAGGAGTACAGGAGGAGTCCGCAGGACG--CCACCCCCACGCAGAGAAAGAAGTCAGTA 473

Qy    301 cccagaaatccatcttccttggaaccaggggtctgcacatccgatgaagatccagaagtgtcag 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    474 CCCAGAAATCATCTTCCTTGGAACCAGGGGTCTGCCATCCGGATGAAGAATCGAAATGTCAG 533

Qy    361 tgccacacttgtcaacataagcccgacacacgtcctctgctactggactgtgtgggggcac 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    534 TGCCACACTTGTCAACATAAGCCCCGACACACGTCTTCTGCTACTGGACTGTGGTGGGGCAC 593

Qy    421 atttgggcaagctgtgccgtcattac 445
          |||||||||||||||||||||||||||||
Db    594 ATTTGGGCAGCTGTGTCCGTCATTAC 618
```

LOCUS   BE250309   574 bp   mRNA   EST   13-JUL-2000
DEFINITION   600943455F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960077 5',
             mRNA sequence.
ACCESSION   BE250309
VERSION   BE250309.1 GI:9120418
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 574)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM52 row: n column: 14
             High quality sequence stop: 571.
FEATURES   Location/Qualifiers
     source   1..574
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2960077"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT   124 a   158 c   167 g   119 t   6 others
ORIGIN

Query Match   31.7%; Score 317; DB 165; Length 574;
Best Local Similarity   99.7%; Pred. No. 2e-157;
Matches 367; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy    503 gcacgcagatcaccacacgggcttgccaagtatcttggctgcaggagagagaacgcgccttggc 562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 GCACGCAGATCACCACACACGGGGCTTGCCAAGTATCTTGCTGCGAGGAGAGAACGCGCCTTGGC 127

Qy    563 atcttgggaaagccgcttcaccccttgctggtgttgttgccccccaacccagctcaaagcctg 622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 16

LOCUS       AU124671

DEFINITION  AU124671      762 bp    mRNA                    EST      23-OCT-2000
            AU124671 NT2RM4 Homo sapiens cDNA clone NT2RM4000375 5', mRNA
            sequence.
ACCESSION   AU124671
VERSION     AU124671.1  GI:10949387
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3951
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RM4000375"
                     /clone_lib="NT2RM4"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector:pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
BASE COUNT      180 a    212 c    202 g    165 t    3 others
ORIGIN

Query Match                31.3%;  Score 313;  DB 107;  Length 762;
Best Local Similarity    99.7%;  Pred. No. 2.7e-155;
Matches   363;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy       1 tcgcagccacaagagattcaaaccacctcaacctcatccacccggacaatcttccccctgct  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 ATCTTTGGGAAAGCCGCTTCACCTTTGCTGGTTGGTTGCCCCACCAGCTCAAAGCCTG  187

Qy      61 gctccagcagtacaccaccagtgccaggaggtcctgcaccacacacacacacacacacactcctgc  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 GCTCCAGCAGTACCACCACCAGTGCCAGGAGGTCCTGCACCACACATCAGTAGATCAGTTC  247

Qy     121 caaatgccttccaggaaagggggctgagatctccagtcctctgtgagaaagatctgatcagttc  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 CAAATGCCTTCAGGAAAGGGGGCTGAGATCTCCAGTCCTCTGCAGTGGAAAGATTGATCAGTTC  307

Qy     181 gctgttgcgaacatgtgattggaagagagttcaagacctgtctgcggcagcactgcaagca  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     308 GCTGTTGCGAACGTGTGATTGGAAGAGAGTTTCAGACCTGTCTGCGGCACTGCAAGCA  367

Qy     241 tgcgtttggcgtcgctggtgtgcacaacctctggctggaaagtggtctattccgggggacac  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     368 TGCGTTTGGCGTCGCTGGTGTGCACAACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC  427

Qy     301 cccagaaatcatctccttggaacacagggtctgccatcccgatgaagattcgaaatgtcag  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     428 CATGCCCCT  435

Qy     361 catgcccct  870
           |||||||||
Db     863 catgcccct  870
```

```
RESULT 17

LOCUS       BF315158

DEFINITION  BF315158      1497 bp    mRNA                    EST      21-NOV-2000
            601899518F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4128701 5',
            mRNA sequence.
ACCESSION   BF315158
VERSION     BF315158.1  GI:11263379
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1497)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1022  row: c  column: 06
            High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1..1497
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4128701"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      512 a    382 c    430 g    173 t
ORIGIN

Qy     289 TCGCAGCCACAAGAATTCAAACCACCTCAACCTCATCCACCCGGACAATCTTCCCCCTGCT  348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      ...

Qy      61 caccagttcccgtgtaagaaggagggccccacccctcagttgtgcccatggttcagggtga  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     349 CACCAGTTTCCCGTGTAAGAAGGAGGGCCCCACCCCTCAGTTGTGCCCATGGTTCAGGGTGA  408

Qy     121 atgcctcctcagtaccagctccgtcccaggaggggagtggcaggaggggatgtgccattac  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     409 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGGAGTGGCAGGAGGGGATGCCATTATTAC  468

Qy     241 gcaggagtacaggagggtgccagacggacgggccccagcagaggagaaaaaaagtcagta  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     469 TTGCAATCCTGAGGAAATTCATAGTTGAGGCGCGTTCCAACTTCCATCAGAGCGT  528

Qy     529 GCAGGAGTACAGGAGGGTGCCAGACGGACGGGCCCCAGCAGAGGAGAAAAAAAGTCACTA  588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy     301 cccagaaatcatctccttggaacacagggtctgccatcccgatgaagattcgaaatgtcag  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     589 CCCAGAAATCATCTTCCTTGGAACACAGGGTCTCCCATCCCGATGAAGAATTCGAAATGTCAG  648

Qy     361 tgcc  364
           ||||
Db     649 TGCC  652
```

RESULT 18

```
LOCUS       BE902696      735 bp    mRNA            EST      29-SEP-2000
DEFINITION  601677393F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',
            mRNA sequence.
ACCESSION   BE902696
VERSION     BE902696.1  GI:10393148
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM841 row: j column: 23
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1..735
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3959926"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
```

Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```
BASE COUNT    168 a    206 c    203 g    157 t    1 others
ORIGIN
```

Query Match   31.2%; Score 312; DB 147; Length 1497;
Best Local Similarity   99.8%; Pred. No. 9.9e-155;
Matches   432; Conservative   0; Mismatches   1; Indels   1; Gaps   1;

```
Qy     43 ggacatcttcccctgctcaccagttccgctgtaagaagaggggcccacctcagtgt 102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 GGACATCTTCCCCTGCTCACCAGTTTCCGCTGTAAGAAGAGGGGCCCCACCTCAGTGT 170

Qy    103 gcccaatggttcagggtgaatccttcctccataccagtccgtccaggaggaggagtggca 162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 GCCCAATGGTTCAGGGTGAATGCCTCCTCCATACCAGTCCGTCCCAGGAGGAGGAGTGGCA 230

Qy    163 gaggggatgccattattattcttgcaatcctgaggaaattcatagttgaggcgctgcagcttcc 222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 GAGGGGATGCCATTATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCC 290

Qy    223 caacttccagacgagctgcaggagtacaggagagtcgccaggagcccagccagcc 282
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 CAACTTCCAGACGAGCCTGCAGGAGTACAGGAGAGTCGCCAGGAGCCCAGCCAGCC 349

Qy    283 agagaaaagaagtcaatcatcttcttggaacaggggtctgtcatcccgat 342
          ||||||||||||||||||||||||||||||||||||| |||||||||||
Db    350 AGAGAAAAGAAGTCAGTCATCTTCCTTGGAACAGGGGTCTGTCATCCCGAT 409

Qy    343 gaagattcgaaatgtcagtgccacacttgtcaacatcttctttggaacagtctctgctact 402
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    410 GAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTTCTCTGCTTACT 469

Qy    403 ggactgtggttgagggccacattgggcagcgtgtcgccgtcattacgggagaccaggttggacag 462
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    470 GGACTGTGGTGAGGGCACATTGGGCAGCGTGTCGCCGTCATTACGGAGACCAGGTGGACAG 529

Qy    463 ggtcctgggcacc 475
          |||||||||||||
Db    530 GGTCCTCTGGGCACC 542
```

RESULT 19

```
LOCUS       AA310236      431 bp    mRNA            EST      19-APR-1997
DEFINITION  EST181085 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA310236
VERSION     AA310236.1  GI:1962584
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 431)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,
            O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,
            L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
            Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,
            M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     96026280
```

Query Match   30.7%; Score 307; DB 141; Length 735;
Best Local Similarity   100.0%; Pred. No. 4.3e-152;
Matches   307; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    183 gcaatcctgaggaaattcatagttgaggcgctgcagcttcccaacttccagacgagcgtgc 242
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 GCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGACGAGCGTGC 193

Qy    243 aggagtacaggagagtcgccaggagcccagccagccagccagaagaaaagaagtcagtacc 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AGGAGTACAGGAGAGTCGCCAGGAGCCCAGCCAGCCAGCCAGAGAAAAGAAGTCAGTACC 253

Qy    303 cagaaatcatcttccttggaacaggggtctgtccaatcccgatgaagattcgaaatgtcagtg 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 CAGAAATCATCTTCCTTGGAACAGGGGTCTGTCCAATCCCGATGAAGATTCGAAATGTCAGTG 313

Qy    363 ccacacttgtcaacataagccccgacacgtctctgctactggactgtggtggagggcacat 422
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 CCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGGAGGGCACAT 373

Qy    423 ttgggcagcgtgtcgccgtcattacggagaccaggttggacagggtcctggggcacctggctg 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    374 TTGGGCAGCGTGTCGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGCACCCTGGCTG 433

Qy    483 ctgtcgtt 489
          ||||||||
Db    434 CTGTCGTT 440
```

```
COMMENT     Other_ESTs: THC175624
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 301699056
            Fax: 301699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            Information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..431
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):156306"
                     /db_xref="taxon:9606"
                     /clone_lib="Jurkat T-cells V"
                     /cell_type="T-lymphocyte"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       98 a     131 c     117 g      84 t      1 others
ORIGIN

Query Match
Best Local Similarity   28.9%;   Score 289; DB 5; Length 431;
                        100.0%;  Pred. No. 1.6e-142;
Matches 289;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 tcgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgct 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  143 TCGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCT 202

Qy   61 caccagtttccgtctgtaagagaaggagggcccccacccctcagtgtccatggttcaggtga 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  203 CACCAGTTTCCGTCTGTAAGAGAAGGAGGGCCCCCACCCCTCAGTGTGCCATGGTTCAGGTGA 262

Qy  121 atgcctcctcaagtaccagctccgtcccagggagggatggcagaggggatgtgccattac 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  263 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGGAGGGATGTGGCAGAGGGGATGCCCATTAC 322

Qy  181 ttgcaatcctgaggaattcatagttgagcgcgtgcagcttcccaacttccagcagaagcgt 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  323 TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGACCGT 382

Qy  241 gcaggagtacaggaggaggagtgcgcaggacggcccagcccccagcagagaaa 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  383 GCAGGAGTACAGGAGGAGGAGTGCGCAGGACGGCCCAGCCCCCAGCAGAGAA 431
```

```
RESULT 20
BE304720
LOCUS       BE304720     612 bp     mRNA        EST     13-JUL-2000
DEFINITION  601106236F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:3349304 5',
            mRNA sequence.
ACCESSION   BE304720
VERSION     BE304720.1   GI:9176150
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM143 row: h column: 09
            High quality sequence start: 21
            High quality sequence stop: 609.
FEATURES             Location/Qualifiers
     source          1..612
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3349304"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming; Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      143 a     178 c     161 g     130 t
ORIGIN

Query Match
Best Local Similarity   28.9%;   Score 289; DB 166; Length 612;
                        100.0%;  Pred. No. 1.6e-142;
Matches 289;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 tcgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgct 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257 TCGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCT 316

Qy   61 caccagtttccgtctgtaagagaaggagggcccccacccctcagtgtccatggttcaggtga 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  317 CACCAGTTTCCGTCTGTAAGAGAAGGAGGGCCCCCACCCCTCAGTGTGCCATGGTTCAGGTGA 376

Qy  121 atgcctcctcaagtaccagctccgtcccagggagggatggcagaggggatgtgccattac 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  377 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGGAGGGATGTGGCAGAGGGGATGCCCATTAC 436

Qy  181 ttgcaatcctgaggaattcatagttgagcgcgtgcagcttcccaacttccagcagaagcgt 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  437 TTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGACCGT 496

Qy  241 gcaggagtacaggaggaggagtgcgcaggacggcccagcccccagcagagaaa 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  497 GCAGGAGTACAGGAGGAGGAGTGCGCAGGACGGCCCAGCCCCCAGCAGAGAAA 545
```

```
RESULT 21
T34216
LOCUS       T34216     307 bp     mRNA        EST     06-SEP-1995
DEFINITION  EST64346 Human White blood cells Homo sapiens cDNA 5' end similar
            to None, mRNA sequence.
ACCESSION   T34216
VERSION     T34216.1   GI:616314
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 307)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.
            ,Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald
            ,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S., Kelley,J.M.,
            Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
```

```
REFERENCE   1  (bases 1 to 947)
AUTHORS     Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W.,
            Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
            Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
            ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
JOURNAL     Nature 377, 3-174 (1995)
MEDLINE     96026280
COMMENT     Other_ESTs: THC15444
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 301869056
            Fax: 301869423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..307
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):1049473"
                     /db_xref="taxon:9606"
                     /clone_lib="Human White blood cells"
                     /tissue_type="white blood cells"
                     /note="Organ: blood"
BASE COUNT       65 a     81 c     86 g     75 t
ORIGIN

Query Match     28.8%;  Score 282; DB 188; Length 307;
Best Local Similarity  100.0%;  Pred. No. 8.1e-139;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  547 gagaacgcgccttggcatctcttgggaaagccgccttcaccccttttgctggttgccccca 606
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCCTTCACCCCTTTGCTGGTTGCCCCCA 60

Qy  607 accagctcaaagcctgggtccagcagtaccacacaccacagtgccaggaggtcctgcaccaca 666
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCAGCTCAAAGCCTGGGTCCAGCAGTACCACAACCACAGTGCCAGGAGGTCCTGCACCACA 120

Qy  667 tcagtatgtgattcctgcaaaatgcccttcaaggagggggctgagatctccagtgtcctgcagtgg 726
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TCAGTATGATATTCCTGCCAAAATGCCCTTCAGGAGGGGGCTGAGATCTCCAGTCCTCCAGTGG 180

Qy  727 aaagattgatcagttcgctgttgcgaacatgttgatttggaaagagtttcagaacctgtctgg 786
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAAGATTGATCAGTTCGCTGTTGCCGTGTTGGAAGAGAGTTTCAGAACCTGTCTGG 240

Qy  787 tgcggcactgcaagcatcggcgttggctcgtcgcgctggtgcaca 828
        ||||||||||||||||||||||||||||||||||||||||||||
Db  241 TGCGGCACTGCAAGCCATGCGTTTGGCTCGTCGCGCTGGTGCACA 282
```

```
COMMENT     Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9783 row: a  column: 06
            High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1..947
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4212365"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.57 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT     224 a    257 c    289 g    177 t
ORIGIN

Query Match     28.0%;  Score 280; DB 150; Length 947;
Best Local Similarity  99.6%;  Pred. No. 1e-137;
Matches 520; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy  241 gcaggagtacaggaggagtcgcaggacgcgccagccccagcagagaaaagaagtcagta 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCAGGAGTACAGGAGGAGTCGCAGGACGCGCCCAGCCCAGCAGAGAAAAGAAGTCAGTA 60

Qy  301 cccagaaaatcatctcccttggaacacagggtctgccatcccgatgaagattcgaaatgtcag 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCCAGAAAATCATCTTCCTTGGAACACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAG 120

Qy  361 tgccacacttgtcaacatagcccgacacgtctctctgctactggactgttggtgagggcac 420
        |||||||||||||| ||||| |||||||| |||||||||||||||||||||||||||||||
Db  121 TGCCACACTTGTCACACATAAGCCCCGACACAGTCTCTGCTACTGGACTGTTGGTGAGGGCAC 180

Qy  421 atttggcagctgtgccgtcattacgagaccaggtggacagggtcctgtgggcaccctggc 480
        |||||| ||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db  181 ATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGC 240

Qy  481 tgctgtgttgttgtcccacctggcacgcagaatcacacacagggcttgccaagtatcttgc 540
        ||||||| |||| |||||||| |||||||| |||||||||||||||||||||||||||||
Db  241 TGCTGTG-TTGTGTCCCACCT-GCACGCAGATCACACACAGGGCTTGCCAAGTATCTTGC 298

Qy  541 tgcagagagaacgcgccttggcatctcttgggaaagccgcttcaacccttgctggttg 600
        ||||||||||||||||||||||||| |||||||||||||||||||  ||||||||||||
Db  299 TGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTTG 358

Qy  601 ccccccaaccagcctcaaagcctggctccagcagtacacacaccacagtgccaggaggtcctgc 660
        |||||| |||||| ||||||||||||||||||||||| |||| ||||||||||||||||||||
Db  359 CCCCCAAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCACAGTGCCAGGAGGTCCTGC 418

Qy  661 accacaatcagtatgatcctgccaaatgcctttcaggaaggggctgagaatccagtcctg 720
        |||||| |||||||||| ||||||||||||||||||||||||||||||| |||||||||
Db  419 ACCACAATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGAATCTCCAGTCCTG 478

Qy  721 cagtggaaaagaattgatcagttcgctgttgcgaaccatgtgatt 762
        |||||||||||||||||||||||||||||||||||| ||||||||
Db  479 CAGTGGAAAAGAATTGATCAGTTCGCTGTTGCGAAACATGTGATT 520
```

```
RESULT  22
LOCUS      BF525432      947 bp    mRNA           EST       11-DEC-2000
DEFINITION 602069517F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4212365
           5', mRNA sequence.
ACCESSION  BF525432
VERSION    BF525432.1  GI:11612793
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 947)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)

RESULT  23
LOCUS      BF342802     1124 bp    mRNA           EST       22-NOV-2000
DEFINITION 602015077F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4150771
```

ACCESSION   BF342802
VERSION   BF342802.1 GI:11289829
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1124)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9414 row: j column: 20
            High quality sequence stop: 670.
FEATURES   Location/Qualifiers
     source   1..1124
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:415071"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.57 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   322 a   281 c   346 g   175 t
ORIGIN

Query Match 27.9%; Score 279; DB 147; Length 1124;
Best Local Similarity 100.0%; Pred. No. 3.6e-137;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   625 tccagcagtaccacaccaccagtcgccagaggtcctgcacccactcagtatgatgattccacca 684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 TCCAGCAGTACCACACCACCAGTCGCCAGGAGGTCCTGCACCCACATCAGTATGATGATTCCCA 65

Qy   685 aatgcctccaggaagggggctgagatctccagtcctgcagtggaaatgttgatcagttcgc 744
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 AATGCCTTCAGGAAGGGGGCTGAGATCTCCCAGTCCTGCAGTGGAAAAGATTGATCAGTTCGC 125

Qy   745 tgttgggaacatgtgattgatttggaagaagtttcagacccgtcgtccggccgcactgcaacgcatg 804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 TGTTGCCAACATGTGATTGATTTGGAAGAGTTTCAGACCCGTCTGGTGCGGCCACTGCAAGCATG 185

Qy   805 cgtttggcttgtcgcgtgttgcacacctctggctgaaagtggtctattccgggacacca 864
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186 CGTTTGGCTTGTCGCGTCGTGCGCGTGGTGCACACCTCTGCCTGAAAGTGGTCTATTCCGGGACACCA 245

Qy   865 tgccctgccgaggcctctcgtccggatgggggaaagatgcca 903
          ||||||||||||||||||||||||||||||||||||||||||
Db   246 TGCCCCTGCCGAGGCCTCTCGTCCGGATGGGGAAAGATGCCA 284

RESULT 24
LOCUS   BE742908   677 bp   mRNA   EST   15-SEP-2000
DEFINITION 601574605F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3835658 5',
            mRNA sequence.
ACCESSION   BE742908
VERSION   BE742908.1 GI:10156900
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 677)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM518 row: a column: 03
            High quality sequence stop: 672.
FEATURES   Location/Qualifiers
     source   1..677
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3835658"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT   160 a   173 c   217 g   127 t
ORIGIN

Query Match 27.1%; Score 271; DB 139; Length 677;
Best Local Similarity 100.0%; Pred. No. 6.3e-133;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   669 agtatgattcctgccaaatgcctcaggaaggggctgagatctccagtcctgcagtggaa 728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGTATGATTCCTGCCAAAATGCCTTCAGGAAGGGGGCTGAGATCTCCAGTCCTGCAGTGGAA 60

Qy   729 agattgatcagttcgctgttgggaacatgtgattgatttggaagaagtttcagacccgtcgtctg 788
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGATTGATCAGTTCGCTGTTGCGAACATGTGATTGATTTGGAAGAGTTTCAGACCCGTCTGGTG 120

Qy   789 cggcactgcaagcatgcgtttggcttgtcgcgtgttgtgcacacctctggctgaaagtggtc 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGGCCACTGCAAGCATGCGTTTGGCTTGTCGCGTCGTGCGCGTGGTGCACACCTCTGGCCTGAAAGTGGTC 180

Qy   849 tattccgggggaccaccatgccctgccgaggcctctcgtccggatgggggaaagatgccaccacctc 908
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TATTCCGGGGACACCATGCCCCTGCCGAGGCCTCTCGTCCGGATGGGGAAAGATGCCACCCTC 240

Qy   909 ctgatacatgaagccaccccctggaagatggtt 939
          |||||||||||||||||||||||||||||||||
Db   241 CTGATACATGAAGCCACCCCTGGAAGATGGTT 271

RESULT 25
LOCUS   R55841   482 bp   mRNA   EST   23-MAY-1995
DEFINITION yg89d01.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:40931 5' similar to SP:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD
            PROTEIN IN SIS2-MTD1 INTERGENIC ;, mRNA sequence.
ACCESSION   R55841
VERSION   R55841.1 GI:825947
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens

```
RESULT 26
LOCUS       BF969043                944 bp    mRNA              EST    22-JAN-2001
DEFINITION  602269966F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357985 5',
            mRNA sequence.
ACCESSION   BF969043
VERSION     BF969043.1  GI:12336258
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9995  row: d  column: 18
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1..944
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4357985"
                     /clone_lib="NIH_MGC_84"
                     /tissue_type="adrenal cortex carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
                     primed; Average insert size 1.229 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      200 a     266 c    264 g   214 t
ORIGIN

Query Match     25.3%;  Score 253;  DB 172;  Length 944;
Best Local Similarity  100.0%;  Pred. No. 2.5e-123;
Matches  253;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 tcgcagccacaagattcaaaacccagtcaacctcatccaccggacatcttccccctgct 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       345 TCGCAGCCACAAGATTCAAAACCCAGTCAACCTCATCCACCGGACATCTTCCCCCTGCT 404

Qy        61 caccagtttccgctgtaagaaggagggccccaccccctcagtgtgccatggttcaggtga 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       405 CACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCCCTCAGTGTGCCATGGTTCAGGTGA 464

Qy       121 atgcctcctcaagtaccagctccgtccccaggagggagggtggcgcagaggagggatgccattattac 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       465 ATGCCTCCTCAAGTACCAGCTCCGTCCCCAGGAGGGAGGGTGGCGCAGAGGGATGCCATTATTAC 524

Qy       181 ttgcaatcctgaggaattcatagttgaggcgtgcagcttcccaacttcccagcagacgct 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       525 TTGCAATCCTGAGGAATTCATAGTTGAGGCGTGCAGCTTCCCAACTTCCCAGCAGACGCT 584

Qy       241 gcaggagtacagg 253
             |||||||||||||
Db       585 GCAGGAGTACAGG 597

RESULT 27
AI991599/c
LOCUS       AI991599                517 bp    mRNA              EST    08-MAR-2000
DEFINITION  ws18c04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2497542 3'
            similar to SW:YATA_SCHPO Q10155 HYPOTHETICAL 90.6 KD PROTEIN
            C1D4.10 IN CHROMOSOME I. ;, mRNA sequence.
ACCESSION   AI991599
```

```
                     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 387 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1777  Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1..482
                     /organism="Homo sapiens"
                     /db_xref="GDB:413472"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:40931"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAGAATTCGGCCGCCGCAGGAATTTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and Hind III and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      107 a     139 c    131 g    102 t      3 others
ORIGIN

Query Match      26.2%;  Score 262;  DB 187;  Length 482;
Best Local Similarity  100.0%;  Pred. No. 3.8e-128;
Matches  262;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       241 gcaggagtacagaggaggagtgcgcagagggacggcccccagcagcagagaaaaagagaaagtcaagta 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        43 GCAGGAGTACAGAGGAGGAGTGCGCGAGAGGGACGGCCCCCAGCAGCAGAGAAAAAGAAGTCAGTA 102

Qy       301 cccagaaatcatctccttggaacacagggtctgtcatcccgatgaagaattcgaaatgtcag 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       103 CCCAGAAATCATCTTCCTTGGAACAGGGGTCTGGCATCCCGATGAAGAATTCGAAATGTCAG 162

Qy       361 tgccacacttgtcaacataagcccgacacagctgctcctgctactgactgttggtgagggcac 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       163 TGCCACACTTGTCAACATAAGCCCGACACAGCTTCTGCTACTGGACTGTGGTGAGGGCAC 222

Qy       421 atttggcagctgtgccgtcattacggagaccaggtgtcctgggcaccctggc 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       223 ATTTGGGCAGCTGTGTGCCGTCGTTCATTACGGAGACCAGGTGTCCTGGGCACCCTGGC 282

Qy       481 tgctgtgtttgtgtcccacctg 502
             |||||||||||||||||||||||
Db       283 TGCTGTGTTTGTGTCCCACCTG 304
```

Left column

```
VERSION     AI991599.1   GI:5838504
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 625    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 426.
FEATURES             Location/Qualifiers
     source          1..517
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2497542"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIds 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo."
BASE COUNT       118 a     139 c     123 g     136 t     1 others
ORIGIN
```

```
Query Match            24.7%;    Score 247;  DB 104;  Length 517;
Best Local Similarity  100.0%;   Pred. No. 3.7e-120;
Matches  247; Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    658 tgcaccacacteagtatgattcctgccaaatgccttcaggaaggaaggggctgagatctccagtc 717
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    447 TGCACCACACTCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCGTGAGATCTCCAGTC 388

Qy    718 ctgcagtggaaagattgatcagttcgctgttcgttgacaaatgtgaaagtttcaga 777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 CTGCAGTGGAAAGATTGATCAGTTCGCTGTTCGGACAACATGTGAAAGAGTTTCAGA 328

Qy    778 cctgtctcggttcgggcaactgcaagcatgcgtttggcgtgttgcgtgcacacctctggct 837
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 CCTGTCTCGGTTCGGGCACTGCAAGCATCCGTTTGGCCTCGTGCGTGGTGCACACCTCTGGCT 268

Qy    838 ggaaagtggttctattccgggaccacacatgcccctgcgagagctctggggggggaaag 897
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 GGAAAGTGGTTCTATTCCGGGACCACACATGCCCTGCGAGGGCTCTGGTCCGGATGGGGGAAAG 208

Qy    898 atggccac 904
          ||||||||
Db    207 ATGGCCAC 201
```

```
RESULT  28
AA634909  .
```

Right column

```
LOCUS       AA634909         446 bp    mRNA             EST        21-OCT-1997
DEFINITION  ab27h02.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
            IMAGE:842067 5' similar to SW:YK59_YEAST P36159 HYPOTHETICAL 96.8
            KD PROTEIN IN SIS2-MTD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AA634909
VERSION     AA634909.1   GI:2558123
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:842067"
                     /clone_lib="Stratagene lung (#937210)"
                     /sex="male"
                     /dev_stage="72 years"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
                     dT. normal lung. Average insert size: 1.0 kb; Uni-ZAP XR
                     Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'."
BASE COUNT       105 a     112 c     131 g      97 t     1 others
ORIGIN
```

```
Query Match            24.6%;    Score 246;  DB 10;   Length 446;
Best Local Similarity  100.0%;   Pred. No. 1.3e-119;
Matches  246; Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    756 tgtgatttggaaggagagttcagacctgtctcggtgtggtcgcggcactgcaagcatggcgtttggctgt 815
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59 TGTGATTTGGAAGAGAGTTTCAGACCTGTCTCGGTGTGGTCGCGGCACTGCAAGCATCCGTTTGGCTGT 118

Qy    816 gcgctggtgcacacctctggctggaaagtgggtctattccgggggaccacacatggccctgcgag 875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 GCGCTGGTGCACACCTCTGGCTGGAAAGTGGGTCTTATTCCGGGGGACCACACATGGCCCTGCGAG 178

Qy    876 gctctggttccggatggggggaaagtgccaccctcctgatacatgaagccaccctggaagat 935
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 GCTCTGGTTCCGGATGGGGGAAAGTGCCACCCTCCTGATACATGAAGCCACCCTGGAAGAT 238

Qy    936 ggtttggaagaggaagcagtggaaaaagacacaagcacacacgtcccaagccatcaagctg 995
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 GGTTTGGAAGAGGAAGCAGTGGAAAAAGACACAAGCACACACGTCCCAAGCCATCAAGCTG 298

Qy    996 gggatg 1001
          ||||||
Db    299 GGGATG 304
```

```
RESULT  29
AW889463/c
```

```
LOCUS       AW89463        346 bp    mRNA                   EST       24-MAY-2000
DEFINITION  RC6-NT0029-240400-011-B07 NT0029 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW89463
VERSION     AW89463.1  GI:8053668
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 346)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE     20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti-&t2-RC6-NT0029-240
            400-011-B07&t3-2000-04-24&t4-1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 310.
FEATURES             Location/Qualifiers
     source          1..346
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0029"
                     /note="Organ: nervous,tumor; Vector: Puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
                     /dev_stage="Adult"
BASE COUNT        75 a      103 c       85 g       83 t
ORIGIN

Query Match             24.3%;  Score 243;  DB 122;  Length 346;
Best Local Similarity   100.0%;  Pred. No. 4.8e-118;
Matches  243;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        641 ccagtgccaggaggtcctgcacacatcagtatgatccttgccaaatgccttcaggaagg 700
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        335 CCAGTGCCAGGAGGTCCTGCACACATCAGTATGATATCCTTGCCAAATGCCTTCAGGAAGG 276

Qy        701 ggcctgagatcttccagtcctgcagtgaaagattgatcaagttcgcgtgttcgaacatgtga 760
              ||||||||||||||||||||||| ||| |||||| || |||| |||||||||||||||||||
Db        275 GGCTGAGATCTTCCAGTCCTGCACTGGAAAGATGTGATCAGTTCGCGTGTTCGAACATGTGA 216

Qy        761 tttggaagagttcagaccgtctgtggtcgcggacactgcaagcatgcgtttggctgcggct 820
              |||||||||||||||||||| ||||||| |||||||||||||||||||||||||||| |||
Db        215 TTTTGGAAGAGTTTCAGACCTGTCTGGTGTCGGCACTGCAAGCATGCGTTTGGCTGCTGGCT 156

Qy        821 ggtgcacacctctggctggaaaagtggttctattccgggacacatttccctgccggagctct 880
              |||||||||||||||||||||||| |||| ||||||||||| ||| | ||||| || ||||
Db        155 GGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACACACCATGCCCTGCGAGGCTCT 96

Qy        881 ggt 883
              |||
Db         95 GGT 93
```

```
RESULT 30
AA311855
LOCUS       AA311855       501 bp    mRNA                   EST       19-APR-1997
DEFINITION  EST182568 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA
            sequence.
ACCESSION   AA311855
VERSION     AA311855.1  GI:1964184
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 501)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., White
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., Bult
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-H., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     96026280
COMMENT     Other_ESTs: THC175624
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlave@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..501
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /db_xref="ATCC (inhost):159063"
                     /clone_lib="Jurkat T-cells VI"
                     /cell_type="T-lymphocyte"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       125 a      148 c      125 g       98 t        5 others
ORIGIN

Query Match             24.0%;  Score 240;  DB 5;  Length 501;
Best Local Similarity   100.0%;  Pred. No. 2e-116;
Matches  240;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 tcgcagccacaagattcaaaccagtcaaacctcatccaaccggaactcttccccctgct 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 TCGCAGCCACAAGATTCAAACCAGTCAAACCTCATCCAACCGGAACTCTTCCCCCTGCT 169

Qy         61 caccagtttccgctgtaagaaggaggcccaccacccttcaagtgtgccaagttcaggggtga 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT  31

LOCUS       BE795434                698 bp    mRNA              EST        20-SEP-2000
DEFINITION  601592991F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946774 5',
            mRNA sequence.
ACCESSION   BE795434
VERSION     BE795434.1  GI:10216632
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   NIH-MGC  http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCID/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM807 row: f column: 23
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1..698
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3946774"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      201 a    212 c    175 g    110 t
ORIGIN

Query Match             23.0%;  Score 230;  DB 140;  Length 698;
Best Local Similarity   100.0%;
Matches  230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        170 CACCAGTTTCCGCTGTAAGAGAGGAGGAGGCCCACCTCAGTGTGCCATGGTTCAGGGTGA 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        121 atgccctcctcaagtaccagctccgtcccagaggaggaggatggcagtggcagtggagat 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 ATGCCCTCCTCAAGTACCAGCTCCGTCCCAGAGGAGGAGGATGGCCATTATTAC 289

Qy        181 ttgcaaatcctgaggaattcataagttgaggggggctgcagcttcccacagagagcgt 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 TTGCAAATCCTGAGGAATTCATAAGTTGAGGGCGCCTGCAGCTTCCCAACTTCCAGCAGGAGCGT 349
```

```
RESULT  32

LOCUS       BE260626                938 bp    mRNA              EST        26-OCT-2000
DEFINITION  601461161F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161691 5',
            mRNA sequence.
ACCESSION   BE260626
VERSION     BE260626.1  GI:9132065
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   NIH-MGC  http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM119 row: o column: 04
            High quality sequence stop: 621.
FEATURES             Location/Qualifiers
     source          1..938
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3161691"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      190 a    263 c    283 g    202 t
ORIGIN

Query Match             22.3%;  Score 223;  DB 165;  Length 938;
Best Local Similarity   100.0%;
Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        683 caaatgccttcaggaagagggctgagatctccagtcctgcagtgggaaagat 732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CAAATGCCTTCAGGAAGAGGGGCTTGAGAGATCTCCAGTCCTTGCAGTGGAAAGAT 392

Qy          1 tcgcagccacaagattcaaaccagctcaacctcacctcaccccgagaatcttcccctgct 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TCGCAGCCACAAGAGTTCAAACCAGCTCAAACCTCAATCCACCCGGACAACTCTTCCCCTGCT 418

Qy         61 caccagtttccgctgtaagaggaggaggcccccaacctcaagttgtcccatggttcaggtga 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 CACCAGTTTCCGCTGTAGAAGGAGGGGGCCCCACCCTCAAGTGTGCCCATGGTTCAGGGTGA 478

Qy        121 atgcctcctcaagtaccagctccgtcccaggaggaggaggatggcagagggatgccattac 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGAGGAGGATGGCCATTATTAC 538

Qy        181 ttgcaatcctgaggagattcatagttgaggcgctgcagcttccc 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 TTGCAATCCTGAGGAAATTCATACAGTTTGAGGCGCTGCACTTCCC 581
```

```
Qy        503 gcacgcagaatccacacacggcttgccaagtatctcgtgcagagagaagacggcccttggc 562
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 GCACGCAGGATCACACAGCGGCTTGCCAAGTATCTTGCGAGAGAGAAACCGGCCCTTGGC 222

Qy        563 atcttggggaaagccgcgcttcacccctttgctggtggttgtgccccccaaccaagcctg 622
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 ATCTTTGGGAAAGCCCCTTCACCCTTTGCTCGTGGTTTGCCCCCAACCAGCTCAAAGCCTG 282

Qy        623 gctccagcagtaccacaaccagtgccaagaggaggtcctgcaccacacatcagtgatgaattcctgc 682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 GCTCCAGCAGTACCACAACCAGTGCCAGGAGGAGGTCCTGCACCACCATCAGTATGAATTCCTGC 342
```

RESULT 33

```
LOCUS       AW806551                  278 bp    mRNA            EST       17-MAY-2000
DEFINITION  IL0-ST00002-160599-003 ST0002 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW806551
VERSION     AW806551.1  GI:7899550
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 278)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
  COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2-IL0-ST0002-160
            599-003&t3-1999-05-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 278.
FEATURES             Location/Qualifiers
     source          1..278
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ST0002"
                     /dev_stage="Adult"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       64 a     84 c     70 g     60 t
ORIGIN
```

Query Match   20.9%;   Score 209;   DB 120;   Length 278;
Best Local Similarity   100.0%;   Pred. No. 6.1e-100;
Matches 209;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 tcgcagccacaagatcaaacccagctcaacctcatccaccggacatctccccctgct     60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 TCGCAGCCACAAGATCAAACCCAGCTCAACCTCATCCACCGGACATCTCCCCCTGCT    129

Qy     61 caccagttccgctgtaagagaaggagggccccaccccctcagtgtgtgccatggggtga   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 CACCAGTTCCGCTGTAAGAGAAGGAGGGCCCCACCCCCTCAGTGTGTGCCATGGGGTGA    189

Qy    121 atgcctcctcaagtaccagctccgtcccaggagggaggggatgccattattac         180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGGGGATGCCATTATTAC       249

Qy    181 ttgcaatcctgaggaattcatagttgagg         209
          |||||||||||||||||||||||||||||||
Db    250 TTGCAATCCTGAGGAATTCATAGTTGAGG       278
```

RESULT 34

```
LOCUS       T34024                    282 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST61387 Human White blood cells Homo sapiens cDNA 5' end similar
            to None, mRNA sequence.
ACCESSION   T34024
VERSION     T34024.1  GI:616122
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M-W., Clayton,R.A.,
            Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald
            ,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,Jr,P.S., Kelley,J.M.,
            Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
            Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W.,
            Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
            Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
            ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
  COMMENT   Other ESTs: THC15444
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 301699056
            Fax: 301699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..282
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):104694"
                     /db_xref="taxon:9606"
                     /clone_lib="Human White blood cells"
                     /clone_lib="Human White blood cells"
                     /tissue_type="white blood cells"
                     /note="Organ: blood"
BASE COUNT       62 a     83 c     73 g     62 t     2 others
ORIGIN
```

Query Match   20.9%;   Score 209;   DB 188;   Length 282;
Best Local Similarity   100.0%;   Pred. No. 6.1e-100;
Matches 209;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    503 gcacgcagatcaccacacggggcgttgccaagtatcttcgtgcaggagaagaacgccctggc    562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 GCACGCAGATCACCACACGGGGCGTTGCCAAGTATCTTCGTGCAGGAGAAGAACGCCCTGGC    104

Qy    563 atcttgggaaagccgcttcaccctttgctgtggttgcccccaaccagctcaaagcctg    622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 ATCTTGGGAAAGCCGCTTCACCCTTTGCTGTGGTTGCCCCCAACCAGCTCAAAGCCTG    164

Qy    623 gctcccagcagtaccacacaaccagtgccagggaggtcctgcaccaacccaatcagtatgatcctgc    682
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 GCTCCAGCAGTACCACACAACCAGTGCCAGGGAGGTCCTGCACCAACCAACATCAGTATGATTCCTGC    224
```

Qy   683 caaatgccttcaggaagggggctgaagtct 711
          ||||||||||||||||||||||||||||||
Db   225 CAAATGCCTTCAGGAAGGGGCTGAGATCT 253

RESULT 35
BG480926
LOCUS   BG480926   963 bp   mRNA   EST   21-MAR-2001
DEFINITION 602530075F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4653330 5',
            mRNA sequence.
ACCESSION   BG480926
VERSION   BG480926.1 GI:13413205
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 963)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1439 row: f column: 19
            High quality sequence stop: 658.
FEATURES   Location/Qualifiers
     source   1..963
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4653330"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT   270 a   246 c   291 g   155 t   1 others
ORIGIN

Best Local Similarity   100.0%;
Matches   206; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Query Match   20.6%;   Score 206;   DB 154;   Length 963;

Qy   734 gatcagttccgtgttgcgaacatgtgatttggaagagtttcagacacctgtctgtgtcggca 793
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 GATCAGTTCCGTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACACCTGTCTGTGTCGGCA 61

Qy   794 ctgcaagcatgcgtttggcgtgctggtgcacacctctggctggaagtggtcttattc 853
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 CTGCAAGCATGCGTTTGGCGTGCTGGTGCACACCTCTGGCTGGAAGTGGTCTTATTC 121

Qy   854 cggggacacacatgccctgcgcgagggctcctgttccggatgggaaagatgccacctcctgat 913
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CGGGGACACACATGCCCTGCGCGAGGGCTCCTGGTCCGGATGGGAAAGATGCCACCTCCTGAT 181

Qy   914 acatgaagccaccctggaagatggtt 939
          ||||||||||||||||||||||||||
Db   182 ACATGAAGCCACCCTGGAAGATGGTT 207

.

Db   225 CAAATGCCTTCAGGAAGGGGCTGAGATCT 253

RESULT 36
R51138
LOCUS   R51138   472 bp   mRNA   EST   18-MAY-1995
DEFINITION yg71c08.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:38752 5' similar to SP:YK59_YEAST P36159 HYPOTHETICAL 96.8 KD
            PROTEIN IN SIS2-MTD1 INTERGENIC ;, mRNA sequence.
ACCESSION   R51138
VERSION   R51138.1 GI:813040
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 472)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE   The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 327
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13Rl
            High quality sequence stop: 327.
FEATURES   Location/Qualifiers
     source   1..472
                     /organism="Homo sapiens"
                     /db_xref="GDB:411293"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:38752"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer (5'
                     AACTGGAAGAATTCGCGGCCGCGCCGAGGAATTTTTTTTTTTTTTTTTT 3');
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT   105 a   131 c   125 g   102 t   9 others
ORIGIN

Query Match   19.9%;   Score 199;   DB 187;   Length 472;
Best Local Similarity   100.0%;
Matches   199; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   241 gcaggagtacaggagggagtgcccagacacggcccagacccagacccagagagaaagaagtcagta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43 GCAGGAGTACAGGAGGAGTGCCGCAGGACGCCCAGACCCCAGACAGCAGAGAAAAGAAGTCAGTA 102

Qy   301 cccagaaatcatcttcctggaacagggtctgccatccgatgaaagatccgaaatgtcag 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   103 CCCAGAAATCATCTTCCTTGGAACAGGGTGTCTCCCATCCCGATGAAGATTGCGAAATGTCAG 162

Qy   361 tgccacacttgtcaacataagccccgacacacgtctctgctactgggactgtggtggggcac 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   163 TGCCACACTTGTCAACATAAGCCCCGACACACGTCTCTGCTACTGGGACTGTGGTTGAGGGCAC 222

Qy   421 atttgggcaagctgtgccgt 439

```
RESULT 37

Db    223  ATTTTGGGCAGCTGTGCCGT 241
           ||||||||||||||||||||
           ||||||||||||||||||||

LOCUS       BG258460          891 bp    mRNA              EST       13-FEB-2001
DEFINITION  6023799938F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4510616 5',
            mRNA sequence.
ACCESSION   BG258460
VERSION     BG258460.1  GI:12768189
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 891)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM0392 row: 1 column: 09
            High quality sequence stop: 710.
FEATURES             Location/Qualifiers
     source          1..891
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4510616"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies."
                     Note: this is a NIH_MGC Library."
BASE COUNT      203 a    246 c    271 g    171 t
ORIGIN

Query Match      19.7%;  Score 197;  DB 175;  Length 891;
Best Local Similarity 100.0%;  Pred. No. 1.7e-93;
Matches  197;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  tcgcagccacaagagttcaaacccagctcaacctcatccaacccggacactcttcccctgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367  TCGCAGCCACACAGAGTTCAAACCCAGCTCAAACCTCATCCACCCCGGACACTCTTCCCCTGCT 426

Qy     61  caccagttccgcgctgtaagagaaggagggcccacaccctcagtgtgccatggttcagggtga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427  CACCAGTTTCCGCGCTGTAAGAAGGAGGGCCCACACCCTCAGTGTGCCCATGGTTCAGGGTGA 486

Qy    121  atgcctcctcagtacagctccgtcccagaggagggagtgcagagggatgccattattac 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487  ATGCCTCCTCCAAGTACCAGCTCCGTCCCAGGAGGAGGGAGTGGCAGAGGGATGCCATTATTAC 546

Qy    181  ttgcaatcctgaggaat 197
           |||||||||||||||||
Db    547  TTGCAATCCTGAGGAAT 563
```

```
RESULT 38

LOCUS       BG336190          906 bp    mRNA              EST       27-FEB-2001
DEFINITION  6020404980F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542594 5',
            mRNA sequence.
ACCESSION   BG336190
VERSION     BG336190.1  GI:13142628
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 906)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1222 row: p column: 19
            High quality sequence stop: 610.
FEATURES             Location/Qualifiers
     source          1..906
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4542594"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and SuperScript II RT (Life Technologies)."
BASE COUNT      230 a    230 c    302 g    144 t
ORIGIN

Query Match      19.5%;  Score 195;  DB 152;  Length 906;
Best Local Similarity 100.0%;  Pred. No. 1.9e-92;
Matches  195;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    764  ggaagagttcagaccgtctggtgcggccaccgcaagcatggcggttggctgctgcggcggt 823
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     32  GGAAGAGTTTCAGACCGTTCTGGTGCGGCCACTGCAAGCCATGCGGTTGGCTGTTGCGGTGGT 91

Qy    824  gcacacctcggctggaaagtggtctattccggggacacacaatgccctgcaggctctggt 883
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92  GCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAACACCATGCCCTGCGAGGCTCTGGT 151

Qy    884  ccggatggggaaagtgccacccctcctgatacatgaagccacccctggaaagatggttgga 943
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152  CCGGATGGGGAAAGATGCCACCCTCCTGATACATGAAGCCACCCCTGGAAGATGGTTTGGA 211

Qy    944  agaggaagcagtgga 958
           |||||||||||||||
Db    212  AGAGGAAGCAGTGGA 226
```

```
RESULT 39

LOCUS       BG387104          961 bp    mRNA              EST       12-MAR-2001
DEFINITION  6024558131F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4584266 5',
            mRNA sequence.
ACCESSION   BG387104
VERSION     BG387104.1  GI:13280550
KEYWORDS    EST.
SOURCE      human.
```

ORGANIZM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 961)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1309 row: i column: 03
            High quality sequence stop: 752.
FEATURES   Location/Qualifiers
     source   1..961
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4584266"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT   228 a   269 c   262 g   202 t
ORIGIN

Query Match   19.5%;   Score 195; DB 153;   Length 961;
Best Local Similarity 100.0%;   Pred. No. 1.9e-92;
Matches 195; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   63 ccagtttccgtctgtaagagaaggaggggcccccacccctcagtgtgtgccacatggttcaggtgaat 122
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 CCAGTTTCCGTCTGTAAGAGAAGGAGGGGCCCCCACCCTCAGTGTGTGCCACATGGTTCAGGTGAAT 370

Qy   123 gcctcctcaagtacagctccgtcccaggaaggaggtggtcaagaggaaggatggccattattactt 182
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 GCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGAGGATGGCCATTATTACTT 430

Qy   183 gcaatcctgaggaattcatagttgaggcgcgtgcagagcttcccacattccagacgagcgtgc 242
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   431 GCAATCCTGAGGAATTCATAGTTGAGGCGCGTGCAGAGCTTCCCAACTTCCAGACGAGCGTGC 490

Qy   243 aggagtacaggagga 257
           |||||||||||||||
Db   491 AGGAGTACAGGAGGA 505

RESULT 40
AA346268
LOCUS   AA346268   394 bp   mRNA   EST   21-APR-1997
DEFINITION EST52407 Greater omentum IV Homo sapiens cDNA 5' end, mRNA
            sequence.
ACCESSION   AA346268
VERSION   AA346268.1 GI:1998525
KEYWORDS   EST.
SOURCE   human.
ORGANIZM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 394)
AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult

---

,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellgrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
, Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
TITLE   Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE   96026280
COMMENT   Other_ESTs: THC175624
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlavge@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES   Location/Qualifiers
     source   1..394
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="ATCC (inhost):147767"
                     /clone_lib="Greater omentum IV"
                     /note="Organ: omentum; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
                     /dev_stage="adult"
BASE COUNT   81 a   108 c   108 g   96 t   1 others
ORIGIN

Query Match   18.5%;   Score 185; DB 6;   Length 394;
Best Local Similarity 100.0%;   Pred. No. 3.8e-87;
Matches 185; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   503 gcacgcagatcaccacacacggggcttgccaagtatcttgtgtgcaagagagaaacgcgccttgc 562
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 GCACGCAGATCACCACACACGGGGCTTGCCAAGTATCTTGTGTGCAGAGAGAACGCGCCCTTGC 144

Qy   563 atcttttgggaagcgcgcttcacccctttgctggtggttgccccacccaaccagctcaaagcctg 622
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 ATCTTTTGGGAAGCGCGCTTCACCCCTTTGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTG 204

Qy   623 gtcccacgcagtaccacacacaccagtgccaggaggtcctgcaccacccagtatatattcctgc 682
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 GCTCCCAGCAGTACCACACAACCAGTGCCAGGAGGTCCTGCACCACCATCAGTATATGATTCCTGC 264

Qy   683 caaat 687
           |||||
Db   265 CAAAT 269

RESULT 41
BG116283
LOCUS   BG116283   936 bp   mRNA   EST   30-JAN-2001
DEFINITION 602318546F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4418843 5',
            mRNA sequence.
ACCESSION   BG116283
VERSION   BG116283.1 GI:12609789

```
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 936)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10153 row: 1 column: 12
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1..936
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4418843"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      209 a    267 c    265 g    195 t
ORIGIN

RESULT 42
Query Match   18.5%;  Score 185;  DB 174;  Length 936;
Best Local Similarity 100.0%;
Matches  185;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
                                  Pred. No. 4.1e-87;

Qy      32 ctcatccaccccggacatcttccccccgtcctcacacagttccgctgtaagaagaggaggcccc 91
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     390 ctcatccaccccggacatcttccccccgtcctcacacagttccgctgtaagaagaggaggcccc 449

Qy      92 accctcagtgtgcccatggttcaggtgaatgcctcctcaagtcctcagctccgtcccagg 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     450 accctcagtgtgcccatggttcaggtgaatgcctcctcaagtcctcagctccgtcccagg 509

Qy     152 agggagtggcagagggatgccattattacttgcaatcctgaggaattcatagttgaggcg 211
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     510 agggagtggcagagggatgccattattacttgcaatcctgaggaattcatagttgaggcg 569

Qy     212 ctgca 216
           |||||
Db     570 ctgca 574
```

```
LOCUS       BF686235      547 bp   mRNA       EST       22-DEC-2000
DEFINITION  602143587F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4304829 5',
            mRNA sequence.
ACCESSION   BF686235
VERSION     BF686235.1  GI:11971643
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1170 row: m column: 22
            High quality sequence stop: 545.
FEATURES             Location/Qualifiers
     source          1..547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4304829"
                     /clone_lib="NIH_MGC_46"
                     /tissue_type="leiomyosarcoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      127 a    151 c    149 g    120 t
ORIGIN

RESULT 43
Query Match   18.2%;  Score 182;  DB 168;  Length 547;
Best Local Similarity 100.0%;
Matches  182;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
                                  Pred. No. 1.5e-85;

Qy       1 tcgcagccacaagattcaaacccagctcaacctcatccaccccggacaatcttccccctgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 TCGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCCGGACAATCTTCCCCCTGCT 425

Qy      61 caccagttccgctgtaagaagaggaggcccccacccctccagttgtgcccatggttcaggggtga 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 CACCAGTTCCGCTGTAAGAAGAGGAGGCCCCCACCCCTCCAGTGTGTGCCCATGGTTCAGGGGTGA 485

Qy     121 atgcctcctcaagtaccagctccgtcccagagaggagggaagtgcaagagggaatgccattattac 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     486 ATGCCTCCTCAAGTACCAGCTCCGTCCCAGAGAGGAGGGAGTGGCAGAGGGAATGCCATTATTAC 545

Qy     181 tt 182
           ||
Db     546 TT 547
```

```
LOCUS       BE794311      688 bp   mRNA       EST       20-SEP-2000
DEFINITION  601591442F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945538 5',
            mRNA sequence.
ACCESSION   BE794311
VERSION     BE794311.1  GI:10215496
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 688)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTP/DTP
```

```
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM804 row: c column: 11
             High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3945538"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      169 a     190 c     185 g     144 t
ORIGIN

Query Match     17.9%; Score 179; DB 140; Length 688;
Best Local Similarity     100.0%; Pred. No. 6.2e-84;
Matches  179; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   94 cctcagtgtgtgccatgtgttcaggtgaatgcctcctccaagtaccagctccgtccaggag 153
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  463 CCTCAGTGTGTGCCATGGTTCAGGTGAATGCCTCCTCCAAGTACCAGCTCCGTCCAGGAG 522

Qy  154 ggagtggcaaggaggagtggcaccattattacttgcaatcctgaggaattcatagttgaggcgct 213
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  523 GGAGTGGCAAGGAGGAGTGGCACCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCT 582

Qy  214 gcagcttcccaacttccagccagcagccgtcgcaggagtacaaggaggaggacgagcggc 272
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  583 GCAGCTTCCCAACTTCCAGCCAGCAGCCGTCGCAGGAGTACAAGGAGGAGGACGAGCGGCT 641

RESULT 44
LOCUS       BF240253    852 bp    mRNA    EST    14-NOV-2000
DEFINITION  601905776F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4133442 5',
            mRNA sequence.
ACCESSION   BF240253
VERSION     BF240253.1  GI:11154176
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1034 row: h column: 19
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
```

```
     source          1..852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4133442"
                     /clone_lib="NIH_MGC_54"
                     /tissue_type="from chronic myelogenous leukemia"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc).
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      188 a     220 c     270 g     174 t
ORIGIN

Query Match     17.5%; Score 175; DB 146; Length 852;
Best Local Similarity     100.0%; Pred. No. 8.6e-82;
Matches  175; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy  613 tcaaagcctggctccagcagtaccacacacccagtgccaggaggtcctgcaccacatcagta 672
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  103 TCAAAGCCTGGCTCCAGCAGTACCACACACCCAGTGCCAGGAGGTCCTGCACCACATCAGTA 162

Qy  673 tgattcctgccaaaatgccttcaggaaggggctgagatccccagtcctccagtggaaaat 732
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163 TGATTCCTGCCAAAATGCCTTCAGGAAGGGGCTGAGATCCCCAGTCCTCCAGTGGAAAGAT 222

Qy  733 tgatcagttcgcgtcgttgcgaacatgtgattggaagagtttcagaacctgtctcgt 787
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  223 TGATCAGTTCGCGTCGTTGCGAACATGTGATTTGGAAGAGTTTCAGAACCTGTCTCGT 277

RESULT 45
LOCUS       H03317    441 bp    mRNA    EST    20-JUN-1995
DEFINITION  yj47e10.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
            IMAGE:151914 5', mRNA sequence.
ACCESSION   H03317
VERSION     H03317.1  GI:866250
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,
            M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1152
            High quality sequence stops: 317
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1152    Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 317.
FEATURES             Location/Qualifiers
```

```
source          1. .441
                /organism="Homo sapiens"
                /db_xref="GDB:564064"
                /db_xref="taxon:9606"
                /clone="IMAGE:151914"
                /clone_lib="Soares placenta Nb2HP"
                /sex="Female"
                /dev_stage="placenta obtained at birth (full term)"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer [5'
                AACTGGAAGAATTCGGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. Library
                went through one round of normalization. Library
                constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT     96 a     102 c     134 g     102 t     7 others
ORIGIN

Query Match     17.4%;      Score 174;      DB 157;      Length 441;
Best Local Similarity  100.0%;      Pred. No. 2.8e-81;
Matches  174;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    757  gtgatttggaagagtttcaagacctgtctggtgcggcactgcaagacatgcggtttgggctgtg   816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     55  GTGATTTGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGACATGCGGTTTGGGCTGTG   114

Qy    817  cgctggttgcacaccctctgggctggaaagtggtctattccggggacacacatgccctgcgagg   876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    115  CGCTGGTTGCACACCCTCTGGGCTGGAAAGTGGTCTATTCCGGGGACACACATGCCCTGCGAGG   174

Qy    877  ctctggttccggatgggaaagatgccaccctcctgataccatgaagccaccctgg   930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175  CTCTGGTTCCGGATGGGAAAGATGCCACCCTCCTGATACCATGAAGCCACCCTGG   228
```

Search completed: October 28, 2001, 22:09:06
Job time: 2744 sec